

Structure, evolution and expression of the duplicated growth hormone genes of common carp (*Cyprinus carpio*)

DISSERTATION

zur Erlangung des akademischen Grades

Doctor rerum naturalium

(*Dr. rer. nat.*)

im Fach Biologie

eingereicht an der

Mathematisch-Naturwissenschaftlichen Fakultät I

der Humboldt-Universität zu Berlin

von

Dipl. Biologin Asiya Murakaeva

geb. am 09. Mai 1973 in Tashkent, UdSSR

Präsident der Humboldt-Universität zu Berlin

Prof. Dr. Dr. h.c. Christoph Marksches

Dekan der Mathematisch-Naturwissenschaftlichen Fakultät I

Prof. Dr. C. Limberg

Gutachter: 1. Prof. Dr. G. A. Brockmann
 2. Prof. Dr. P. Hammerstein
 3. Prof. Dr. R. Tiedemann

Tag der mündlichen Prüfung 28.11.2008

Berlin 2008

Abstract

Der Karpfen, *Cyprinus carpio*, ist eine tetraploide Fischart aus der Familie Cyprinidae, die vor 20-50 Mio Jahren entstanden ist. Das Ziel der vorliegenden Arbeit war der Versuch, die funktionelle Rolle der duplizierten GH Gene des Karpfens durch das Studium ihrer Struktur, Evolution und Expression zu verstehen. Die Introns des zweiten GH Gens des Karpfens wurden erstmalig sequenziert und Sequenzvergleiche der kodierenden und nicht-kodierenden Bereiche von Allelen beider GH Gene wurden vorgenommen. Eine phylogenetische Analyse wurde durchgeführt, um die Beziehungen der GH Gene des Karpfens zu denen des tetraploiden Goldfischs und anderer diploider Cypriniden zu untersuchen. Zusätzlich wurden weitere duplizierte Gene des Karpfens, von denen einige auch für das Wachstum von Bedeutung sind, phylogenetisch analysiert. Der Test der relativen Evolutionsrate nach Tajima (1993) zeigte einen statistisch signifikanten Anstieg der Evolutionsrate des GH I Gens beim Karpfen. Es wurden in der vorliegenden Arbeit einige weitere duplizierte Genpaare des Karpfens und Goldfischs gefunden, die ebenfalls eine Lockerung funktioneller Zwänge oder sogar Beweise für positive Darwin'sche Selektion bei einem der beiden Duplikate zeigen. Der Expressionstest hat gezeigt, dass die GH I und GH II Gene auf identischen Niveaus bei Karpfenbrut exprimiert werden, während bei ein Jahr alten Karpfen, drei Jahre alten Männchen und Weibchen sowie den 10 Monate alten, an kalte Temperaturen (2°C) angepassten Fischen die Expression von GH II statistisch signifikant geringer war als die von GH I. Es wurde eine neue und einfache Methode zur Herstellung von rekombinanten, biologisch aktiven GH-Proteinen ohne Notwendigkeit des Refolding entwickelt. Sie ermöglicht spätere Tests, ob die Aktivität von unterschiedlichen GH-Varianten des Karpfens gleich oder unterschiedlich ist.

Schlagwörter: Tetraploidisierung des Genoms, Duplizierte Gene, Evolution der duplizierten Gene, Phylogenetische Analyse, Messung des Gene Expression, Rekombinante Proteine.

Abstract

The common carp, *Cyprinus carpio*, is a tetraploid fish species from the family Cyprinidae that arose about 20-50 Myr ago. The aim of the present work was attempting to understand the functional role of the duplicated common carp GH genes by studying their structure, evolution and expression. The introns of the second GH gene of common carp were sequenced for the first time and sequence comparisons of coding and non-coding regions of alleles of both GH genes were carried out. A phylogenetic analysis was done to examine the relationships of common carp GH genes with GH genes of the tetraploid goldfish and other diploid Cyprinids. In addition, phylogenetic analyses were done with other duplicated genes of common carp, some of which also important for growth. The relative rate test of Tajima (1993) showed a statistically significant increase in the evolution rate of the common carp GH I gene. In addition, some other duplicated gene pairs in common carp and goldfish with relaxation of functional constraints or even evidence of positive Darwinian selection in one of the two gene duplicates were found in the present study. The test of expression rates of the two GH genes has shown that the GH I and GH II genes were expressed at similar levels in carp fry. In contrast, the expression of GH II was statistically significantly lower than that of GH I in one year old carp, three years old males and females as well as in 10 months old fish adapted to cold temperature (2°C). To enable testing the hypothesis if activity of GH diverged between different GH variants of common carp a new and simple method for production of recombinant, biologically active GH proteins without the necessity of refolding was developed.

Keywords: Tetraploidization of genome, Duplicated genes, Evolution of duplicated genes, Phylogenetic analysis, Gene expression analysis, Recombinant proteins.

Contents

Chapter I General Introduction	6
Chapter II Sequence variation of the two common carp growth hormone genes	18
II.1 Introduction.....	18
II.2 Material and Methods	20
II.2.1 DNA isolation, PCR, sequencing and cloning conditions.....	20
II.2.2 Comparison of the two common carp <i>GH</i> genes.....	22
II.2.3 Population examination on the extent of polymorphism.....	22
II.3 Results.....	24
II.3.1 Sequencing the introns of <i>GH II</i> and comparison of both <i>GH</i> genes	24
II.3.2 Extent of polymorphism for large insertion/deletion in third intron of <i>GH I</i> and <i>GH II</i>	33
II.4 Discussion.....	36
Chapter III Phylogeny and evolution of duplicated genes of common carp and goldfish	41
III.1 Introduction	41
III.2 Material and Methods.....	43
III.3 Results	44
III.4 Discussion.....	66
Chapter IV Expression of the two common carp growth hormone genes in relation to ontogenesis, sex and water temperature	72
IV.1 Introduction	72
IV.1.1 Growth hormone in growth, development, and maintenance of energy homeostasis	72
IV.1.2 Differences in the timing and/or pattern of expression of duplicated genes.....	73
IV.1.3 Advantage of the real-time PCR technique in examination of gene expression	74
IV.2 Material and Methods.....	78
IV.2.1 Experimental fish	78
IV.2.2 Sample collection and tissue RNA preparation	78
IV.2.3 First strand cDNA synthesis.....	79
IV.2.4 Quantitative RT-PCR and data analysis.....	79

IV.3 Results	82
IV.4 Discussion.....	88
Chapter V Elaboration of a simple, fast and highly productive approach for the production of soluble common carp recombinant growth hormone.....	92
V.1 Introduction.....	92
V.1.1 Studies of functional divergence of duplicated genes	92
V.1.2 Strategies for improving the yield of soluble recombinant proteins in <i>E. coli</i>	93
V.1.3 Strategies for improving the purification of recombinant proteins	95
V.2 Material and Methods	96
V.2.1 Experimental fish and collection of samples	96
V.2.2 RNA preparation and first strand cDNA synthesis.....	96
V.2.3 Primers and PCR conditions for amplification of <i>GH I</i> and <i>GH II</i> cDNA.....	96
V.2.4 Cloning of PCR products from section V.2.3.....	97
V.2.5 Expression vectors and <i>E. coli</i> strains used for improving the yield of native recombinant protein.....	97
V.2.6 Recloning of <i>GH I</i> and <i>GH II</i> coding sequences in expression vectors	99
V.2.7 Preparation of competent cells.....	102
V.2.8 Transformation of the expression vectors with inserts into competent cells.....	102
V.2.9 Cultivation medium	102
V.2.10 Small-scale protein induction	103
V.2.11 Large-scale protein induction	104
V.2.12 Polyacrylamide gel electrophoresis of proteins.....	104
V.2.13 Polyacrylamide gel staining.....	105
V.2.14 Protein concentration determination.....	105
V.2.15 Purification of polyhistidine-containing soluble recombinant protein	106
V.2.16 Determination of the monomer content.....	106
V.2.17 Recombinant common carp GH activity test.....	107
V.3 Results and Discussion	110
Chapter VI Conclusions	119
Bibliography	125
Appendix	144
Abbreviations	178
Acknowledgments	181

Chapter I General Introduction

The common carp, *Cyprinus carpio* L., has a long history of domestication in Asia (China) as well as in Europe, and numerous strains and breeds have been developed from its wild ancestor. The domesticated/cultured forms attained worldwide distribution: today, carp is among the most important species in freshwater fish culture mainly raised for human consumption with an annual production currently exceeding 3.11 million tons (FAO Fishery Statistics, 2005). A fast growth rate combined with a high vitality is the economically most important trait of the species.

The growth hormone (GH) or somatotropin, a 20 to 22 kDa polypeptide produced by the pituitary gland is essential for normal growth and development of vertebrates. It regulates somatic growth and metabolism of nitrogen, lipid, carbohydrate, and mineral (Davidson, 1987; Harvey et al., 1995). Additionally, in fish it is also involved in osmoregulation (Bolton et al., 1987).

The secretion of GH is regulated in a multifactorial manner including external (environmental) as well as internal, genetic and physiological factors (Fig. I.1). All these regulatory effectors are strongly interacting and their influence may vary in time, space, and combination of both. Somatostatin is the primary inhibitor of basal and stimulated growth hormone secretion (Cook & Peter, 1984; Peter, 1986; Marchant et al., 1987; Lin et al., 1993). GH-releasing factor (GRF) is the primary stimulator of GH secretion (Marchant et al., 1989; Vaughan et al., 1992).

The action of growth hormone occurs through binding with the growth hormone receptor. The receptors for growth hormone are present in liver, kidney, skin, muscle, fat, spleen, brain, ovary, testis and gill of teleosts (Gray et al., 1990; Hirano, 1991; Yao et al., 1991; Sakamoto & Hirano, 1991; Harvey et al., 1995).

Several lines of evidence suggest an interaction between GH and insulin like growth factor I (IGF - I) in the regulation of body growth in teleosts. Administration of GH has been shown to stimulate hepatic IGF - I mRNA expression in the coho salmon (Cao et al., 1989; Duan et al., 1993) and Japanese eel (Duan & Hirano, 1992), and to elevate serum IGF - I-like activity as measured by bioassay or heterologous radioimmunoassay in several species (Bern et al., 1991) including the Japanese eel (Duan & Inui, 1990a,b). The hormonal effect in IGF - I mRNA expression appears to be specific to GH in the coho salmon, as other pituitary hormones (prolactin and somatolactin) were found to be much less potent (Duan et al., 1993). Finally, administration of IGF - I has been shown to increase body weight and length in coho salmon (McCormick et al., 1992). This result, in combination with the

finding that hepatic IGF - I mRNA expression and circulating IGF - I like activity is regulated by GH, provides strong evidence that IGF - I mediates at least some of the growth-promoting effects of GH in teleosts. In addition, IGF - I can suppress the secretion of GH from pituitary in human and in teleost that may represent an autoregulatory component in GH regulation (Yamashita et al., 1986; Perez-Sanchez, 1992).

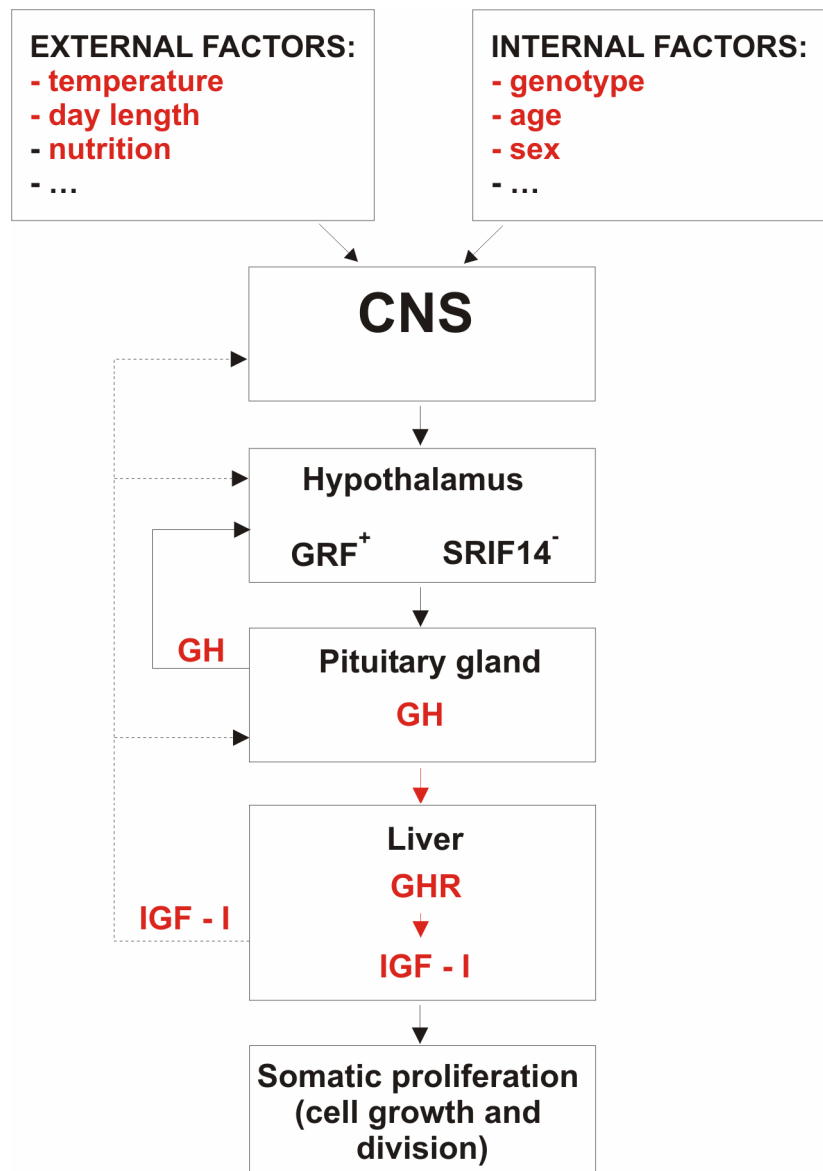


Figure I.1 Schematic presentation of growth hormone regulation and action. CNS - central nervous system, GRF - growth hormone-releasing factor, SRIF14 - somatostatin, GH -growth hormone, GHR - growth hormone receptor, IGF - I - insulin like growth factor. The lines ending in arrows indicate the feedback effect.

The common carp has a very high chromosome number of 98 to 104, approximately twice the number of most other cyprinid fishes (Ohno et al., 1967; Ojima & Hitotsumachi,

1967; Wolf et al., 1969; Raicu et al., 1972; Klinkhardt et al., 1995; Arkhipchuk, 1999). Since the carp also has a high nuclear DNA content, it was proposed to be tetraploid (Ohno & Atkin, 1966). Additional evidence for carp tetraploidy was provided by the finding that many enzymes are expressed in duplicate (Engel et al., 1971; Ferris & Whitt, 1977). Ferris & Whitt (1977) estimated the proportion of still expressed duplicated loci in carp to be about 52%, a proportion similar to that of tetraploid catostomid fishes (suckers). As the catostomids diverged from the cyprinids some 50 Myr ago (Uyeno & Smith, 1972) and as they are considered to have been tetraploid since then (Ferris & Whitt, 1977; Kirpichnikov, 1981), the carp tetraploidy was assumed to be of a similar age (Kirpichnikov, 1981). However, later studies on the basis of flanking regions of duplicated microsatellite loci and sequence analyses of duplicated carp loci have shown that tetraploidization in this fish species took place between 11 to 21 Myr ago (Larhammar & Risinger, 1994; David et al., 2003). Therefore, common carp is one of the vertebrates in which recent genome duplication has occurred.

Since the goldfish, *Carassius auratus*, readily interbreeds with the carp (Hubbs, 1955), these two species probably diverged from a common tetraploid ancestor. The tetraploidization event in goldfish was deduced to occur 16-22 Myr ago, estimated by divergence of the synaptosome-associated loci (Risinger & Larhammar, 1993). Two observations suggested that the tetraploidy resulted from allotetraploidization (species hybridization) rather than from autotetraploidization (genome doubling): no chromosomes seem to have been lost in the tetraploidization event, and no chromosome quadrivalents have been observed in meiotic nuclei (Ohno et al., 1967), indicating that the two original diploid chromosome sets were already different from each other at the time of tetraploidization. Diploid cyprinids have been observed to form hybrids in nature (Hubbs, 1955). Studies of isozyme expression in goldfish (Woods & Buth, 1984) showed that this species still expresses 19% of its genes in duplicate. Therefore, in terms of gene expression, common carp is more similar to the ancestral tetraploid condition than goldfish.

Several salmonid fish species like Atlantic salmon (*Salmo salar* L.), chum salmon (*Oncorhynchus keta* (Walbaum)), rainbow trout (*Oncorhynchus mykiss* (Walbaum)) and sockeye salmon (*Oncorhynchus nerka* (Walbaum)) in which a recent lineage-specific genome duplication occurred were found to have two *GH* genes (Agellon & Chen, 1986; Agellon et al., 1988; Johansen et al., 1989; Lorens et al., 1989; Sekine et al., 1989; Devlin, 1993). Thus, it was supposed that common carp should also possess doubled number of *GH* genes that are expressed. So far, two cDNA sequences (Chao et al., 1989; Koren et al., 1989) and one complete gene sequence (Chiou et al., 1990) had been reported for GH.

Chiou et al. (1990) stated that the genomic clone they had analysed would not give rise to the mRNA described in their previous report (Chao et al., 1989) suggesting the existence of more than one gene. Koren et al. (1989) showed by Southern-blot hybridization with carp *GH* cDNA probes that the carp genome has at least two GH coding regions. Larhammar & Risinger (1994) compared the three published sequences and concluded that those reported by Koren et al. (1989) and Chiou et al. (1990) were most probably alleles of the same locus (*GH I*) whereas the sequence reported by Chao et al. (1989) could represent a second locus (*GH II*). Only 2 silent replacements were found in coding region between the complete *GH* gene sequence of Chiou et al. (1990) and the cDNA sequence of Koren et al. (1989). The differences between the *GH* gene sequence of Chiou et al. (1990) and cDNA sequence of Chao et al. (1989) were 27 substitutions in coding regions, 7 of which resulted in amino acid replacements, and there were 45 substitutions and 19 insertions/deletions in the 3' untranslated region. The introns of the *GH II* gene have not been sequenced up to now.

For a better understanding of the role of each of two growth hormone genes in common carp they should be analysed using knowledge about evolution of duplicated genes. Ohno's influential book *Evolution by gene duplication* (1970) dealt with the idea that gene and genome duplication events are the principal forces by which the genetic raw material is provided for increasing complexity during evolution. He postulated that the major advances in evolution such as the transition from single-celled organisms to complex multicellular animals and plants could not simply have been brought about alone through processes such as natural selection based on existing allelic variation at particular genetic loci in populations. He suggested instead that novelty in evolution is most often based on genomic redundancy as a substrate for subsequent divergent natural selection created initially by gene or entire genome duplications. In a statement that brought his conviction to a point he postulated that “*natural selection merely modified, while redundancy created*”. He meant that gene and genome duplications allowed for gene functions of duplicated genes to diversify, take on novel functions, and bring about evolutionary innovation in general.

Genomic analyses have clarified the extent to which gene duplications occur (Lundin, 1993; Wolfe & Shields, 1997; Postlethwait et al., 1998; Arabidopsis Genome Initiative, 2000; Lynch & Conery, 2000; Vision et al., 2000). These studies have found that a remarkably high fraction of genes are closely related to other genes within the genome. For example, the fraction of genes that represent recent duplication events (recent enough to generate recognizable paralogues) is 11.2% in *Haemophilus influenzae*, 28.6% in *Saccharomyces cerevisiae*, 65.0% in *Arabidopsis thaliana*, 27.5% in *Drosophila melanogaster*, and 44.8% in *Caenorabditis elegans* (Arabidopsis Genome Initiative, 2000).

These numbers increase substantially if less stringent criteria are used to identify paralogues (Rubin et al., 2000).

Ohno (1970) has hypothesized that the increased complexity and genome size of vertebrates has resulted from two rounds (2R) of whole genome duplication (WGD) occurring in early vertebrate evolution that provided the requisite raw materials. Conflicting analyses have now made this very controversial, with some studies supporting 2R (Lundin, 1993; Spring, 1997; Meyer & Schartl, 1999; Lahrhammar et al., 2002) others seeing only a single round of WGD (Guigo et al., 1996; McLysaght et al., 2002), and still others refuting WGD altogether by concluding that nothing greater than limited segmental duplications have occurred (Friedman & Hughes, 2001, 2003).

At the present time polyploids are concentrated in a few vertebrate groups. Except for one rodent, fishes and amphibians represent all known cases of sexual polyploids in vertebrates (Otto & Whitton, 2000). More detailed molecular studies of the evolution of duplicated genes in polyploid species could help to understand if whole genome duplication could conduce to increased complexity of organism. A surprisingly high proportion of genes that duplicated via polyploidization have been retained over long period of time: 72% in maize over 11 Myr (Ahn & Tanksley, 1993; Gaut & Doebley, 1997), 77% in *Xenopus* over 30 Myr (Hughes & Hughes, 1993), 70% in salmonids over 25-100 Myr (Bailey et al., 1978), 47% in catostomids over 50 Myr (Nadeu & Sankoff, 1997). Many of these duplicates have diversified in function and, more typically, in expression patterns.

The studies of duplicated genes in common carp could therefore be useful for better understanding of duplicated gene evolution after genome duplication, the pathways that are involved in the retention and divergence of gene duplicates.

In general, four eventual fates exist for duplicated genes. *I*. One copy is suppressed, either by physical deletion or by accumulation of point mutations until it becomes a pseudogene (pseudogenization). *II*. Both copies may persist in the genome with perfect (or near-perfect) sequence identity (conservation of gene function). Finally, random mutations may cause at least one of the two copies to diverge functionally, either *III* by specializing some aspect of its previous role (subfunctionalization) or *IV* by finding a novel functional role (neofunctionalization).

a Pseudogenization

If the integrity of the two gene copies is not actively maintained, the most likely fate of a gene duplicate is extinction, i.e. inactivation or loss (Haldane, 1933; Bailey et al., 1978; Takahata & Maruyama, 1979; Watterson, 1983; Walsh, 1995). Gene duplication generates

functional redundancy as it is often not advantageous to have two identical genes. In other words, mutations disrupting the structure and function of one of the two genes are not deleterious and are not removed by selection. Gradually, the mutation containing gene becomes a pseudogene, which is either unexpressed or functionless, an evolutionary fate that has been shown by population genetic modelling (Takahata & Maruyama, 1979; Walsh, 1995; Lynch et al., 2001) as well as by genomic analysis (Lynch & Conery, 2000; Harisson et al., 2002). Inactivated gene duplicates (pseudogenes and/or poorly expressed gene copies) have been observed in many gene families (e.g. the MHC class Ib genes; Hughes & Nei, 1989).

Some duplicated genes had been maintained in the genome for a long time for specific functions, before recently becoming pseudogenes because of the relaxation of functional constraints. For example, the size of the olfactory receptor gene family (1000) is similar in humans and mice, but the percentage of pseudogenes is >60% in humans and only 20% in mice. Many olfactory receptor genes have become pseudogenes since the origin of hominoids (Rouquier et al., 2000). This is probably related to the reduced use of olfaction in hominoids, which can be compensated for by other sensory mechanisms, such as better vision.

Pseudogenes do occasionally serve some function. In chickens, there is only one functional gene (VH1) encoding the heavy chain variable region of immunoglobulins, and immunoglobulin diversity is generated by gene conversion of the VH1 gene by the many duplicated variable region pseudogenes that occur on its 5' side (Ota & Nei, 1995).

In Zhang's (2003) view, there have not been sufficient studies of pseudogenization probably because pseudogenes are regarded to be uninteresting. But in fact, lineage-specific pseudogenization, such as the example of olfactory receptor genes of hominoids (Rouquier et al., 2000), provides rich information about organismal evolution. In addition, Lynch & Conery (2000) and Lynch & Force (2000) have recently presented a model suggesting that loss or silencing of duplicated genes might be very important for the evolution of species diversity. They described how the loss of different copies of a duplicated gene in geographically separated populations could genetically isolate these populations, should they become reunited. Genome duplication provides thousands of duplicated genes. Different evolution and silencing of duplicated genes in separated populations could result in fast genetic isolation of these populations.

b Conservation of gene function

Whether two gene copies can be maintained over long periods of evolutionary time depends strongly on whether the genes are redundant in function. If the two gene copies are perfectly redundant, i.e. if a single functional copy can completely mask the deleterious effects of mutations in the other copy, they are very unlikely to be maintained (Fisher, 1935; Cooke et al., 1997). Fisher (1935) showed that with complete redundancy and total overlap in function the maintenance of both copies occurs only if one gene performs its function less efficiently and is less mutable than the other gene. Fisher also noted that both copies could be maintained if they had the exact same efficiency and number of mutations, but this would only work in an infinitely large population. Bailey et al. (1978) also suggested that the large effective population sizes characteristic of salmonids and catostomids permit selection against double null homozygotes to affect gene frequencies. On the other hand, the long-term maintenance of both gene copies is much more likely whenever redundancy is incomplete, such that mutations are deleterious even when they occur in only one gene copy (Takahata & Maruyama, 1979). In this case, selection acts against changes at critical sites in either copy, although the copies would diverge in sequence at silent sites.

Hughes & Hughes (1993) found that amino acid change was rare relative to the rate of silent substitutions in both copies in 17 duplicated pairs of genes in the anciently tetraploid (30 Myr) *Xenopus laevis*. This suggests that neither copy of the gene has been free to accumulate amino-acid altering mutations. In polyploid species a high level of duplicate gene preservation occurs (Ferris & Whitt, 1979; Nadeau & Sankoff, 1997; Amores et al., 1998; Wendel, 2000). It may be reconciled if dosage requirements play an important role in the selective environment of gene duplicates. Polyploidization preserves the necessary stoichiometric relationships between gene products, which may be subsequently maintained by stabilizing selection. On the other hand, duplicates of single genes that are out of balance with their interaction partners may be actively opposed by purifying selection.

Gibson & Spring (1998) have suggested that alteration of a single domain in a multidomain protein might lead to nonfunctional complexes that exhibit a so-called dominant-negative phenotype. Their model is based on the observation that, for several genes, point mutations lead to a much more severe phenotype than when the duplicated gene is simply knocked out. In this case, one would expect selection against deleterious point mutations resulting in the retention of the genes. As a matter of fact, the gene is not only retained, it is also kept redundant.

The presence of duplicate genes is sometimes beneficial simply because extra amounts of protein or RNA products are provided. This applies mainly to strongly expressed genes the products of which are in high demand, such as rRNA and histones. For example, the maintenance of 579 tRNAs within the *C. elegans* genome (Duret, 2000) may be explained by the need for a large pool of tRNAs within the cell, such that loss of tRNA is deleterious even though functioning duplicates of the same tRNA exist within the genome.

c Subfunctionalization

Theoretical population genetics predicates that both duplicates can be stably maintained if they differ in some aspects of their functions (Nowak et al., 1997) that can occur by subfunctionalization, in which each daughter gene adopts part of the functions of their parental gene (Jensen, 1976; Orgel, 1977; Hughes, 1994). Hughes (1994) suggested that the chance of diversification for duplicated genes might be substantially higher whenever the ancestral gene performed multiple functions. He termed it “gene sharing”. In this case, selection might actively favour the specialization of each gene copy on different subsets of these functions. As an example, he cites δ - crystallin, a gene that encodes both an enzyme (argininosuccinate lyase) and an eye-lens crystallin in ducks; in other vertebrates, however, this gene appears to have duplicated such that each gene encodes only one of these products.

One form of subfunctionalization that is potentially important in the evolution of development is division of gene expression after duplication. Force et al. (1999) and Lynch & Force (2000) found that if the original gene had several functional or regulatory domains, deleterious mutations disrupting different subfunctions could accumulate in each of the gene copies. Subsequently, both gene copies would be essential and could not be further inactivated (Werth & Windham, 1991). The fact that duplicate genes often differ in the timing and/or pattern of expression is consistent with this model (Ferris & Whitt, 1979; Hughes & Hughes, 1993; Force et al., 1999). In addition, Force (1999) noted that, with both the engrailed genes, *eng1* and *eng1b*, in zebrafish and the ZAG1 and ZMM2 genes in maize, the shared expression pattern of the duplicated genes matches the total expression pattern of single-copy genes in related organisms lacking the duplication. Zebrafish *engrailed-1* is expressed in the pectoral appendage bud, whereas *engrailed-1b* is expressed in specific set of neurons in the hindbrain/spinal cord (Force et al., 1999). The sole *engrailed-1* gene of the mouse and chicken, orthologous to both genes of zebrafish, is expressed in both pectoral appendage bud and hindbrain/spinal cord (Joyner & Martin, 1987; Davis et al., 1991; Gardner & Barald, 1992).

d Neofunctionalization

One of the most important outcomes of gene duplication is the origin of novel function. Although it seems improbable that an entirely new function could emerge in a duplicate gene, there are several examples. For instance, the eosinophil-derived neurotoxin (EDN) and eosinophil cationic protein (ECP) genes of humans were generated in the lineage of hominoids and Old World monkeys via gene duplication (Zhang et al., 1998). Both genes belong to the RNase A gene superfamily. After duplication, a novel antibacterial activity emerged in ECP. This activity is absent in human EDN and the EDN of New World monkeys, which represents the progenitor gene before duplication. More surprisingly, the antibacterial activity of ECP does not depend on the ribonuclease activity (Rosenberg, 1995). Molecular evolutionary analysis suggests that the new function is probably conferred by a large number of arginine substitutions that occurred in short period after duplication (Zhang et al., 1998). ECP is toxic to bacteria because it makes their cell membranes porous: the positively charged arginine residues might be important for establishing tight contact between the ECP and negatively charged bacterial cell membranes in the pore-formation process (Zhang et al., 1998).

In many cases, however, a related function, rather than an entirely new function, evolves after gene duplication. One good example are the red- and green-sensitive opsin genes of humans, which were generated by gene duplication in hominoids and Old World monkeys (Yokoyama & Yokoyama, 1989). After duplication, the two opsins have diverged in function, resulting in a 30 nm difference in the maximum absorption wavelength. This confers the sensitivity to a wide range of colours that humans and related primates have.

Neofunctionalization of duplicated genes requires varying numbers of amino acid substitutions. The functional change in ECP probably required many substitutions (Zhang et al., 1998) but the functional difference between the red and green opsins is largely attributable to two substitutions (Asenjo et al., 1994).

e Evolutionary forces behind functional divergence of duplicate genes

The most widely cited model of the evolution of functionally novel proteins after gene duplication is based on the assumption that, once gene duplication has occurred, one of the two gene copies is redundant and thus is free from all functional constraints. All mutations occurring in such a redundant gene will be selectively neutral. However, by chance, one or more mutations that preadapt the gene to a new function may occur and may then be fixed by random genetic drift (Ohno, 1973). There are several lines of evidence

inconsistent with the hypothesis: (1) Data from a tetraploid animal, the frog *Xenopus laevis*, suggests that, as long as both copies of duplicate genes are expressed, both are subject to purifying selection (Hughes & Hughes, 1993). Thus, contrary one gene copy is not “ignored by natural selection” and free to mutate at random. (2) A single gene may share more than one function (Piatigorsky & Winstow, 1991) and it was found that functions of the progenitor gene are partitioned between the duplicates so that the joint levels and patterns of activity of the duplicates are equivalent to that of the progenitor gene (Hughes, 1994; Forcer et al., 1999). (3) Several cases of positive selection after gene duplication have been reported, including immunoglobulins, conotoxins, ribonucleases, pregnancy-associated glycoproteins, triosephosphate isomerase and the ECP gene (Tanaka & Nei, 1989; Zhang et al., 1998, 2000, 2002; Hughes, 1999; Duda & Palumbi, 1999; Hughes et al., 2000; Merritt & Quattron, 2001).

In the case of division of expression, such as the *engrailed-1* and *engrailed-1b* genes of zebrafish, it is likely that random fixations of complementary degenerate mutations under relaxed functional constraints are the main cause (Force et al., 1999). In other words, it is a result of neutral evolution, without the involvement of positive selection.

In the case of functional specialization and neofunctionalization the following model was suggested in the last few years. It requires positive selection and involves two scenarios. In the first scenario, after gene duplication, a few neutral or nearly neutral substitutions create a new, but only weakly active function in one daughter gene and positive selection then accelerates the fixation of advantageous mutations that enhance the activity of the novel function (Zhang et al., 1998). In the second scenario, the ancestral gene already has dual functions. Gene duplication provides the opportunity for each daughter gene to adopt one ancestral function, and further substitutions under positive selection can refine the functions (Hughes, 1999).

Acceleration of protein sequence evolution following gene duplication is often observed (Ohta, 1994; Lynch & Conery, 2000; Van de Peer et al., 2001). It is often assumed as neutral evolution with relaxed purifying selection.

Zhang et al. (2002) state that:

“The relaxation of purifying selection is often treated as the null hypothesis and thus accepted even without direct evidence. Because of the relatively low power of statistical methods for detecting positive selection, actions of positive selection have probably been overlooked and relaxation of purifying selection incorrectly invoked. Both positive selection and relaxation of purifying selection are necessary in the functional divergence of duplicate

genes. It might be particularly so when the functional change involves multiple amino acid substitutions.”

He & Zhang (2005) have analysed genome-wide patterns of yeast protein interaction and gene expression for human duplicated genes. Their analysis revealed rapid subfunctionalization of duplicated genes. Degenerative mutations reduce joint levels of duplicated genes and patterns of activity to that of the single ancestral gene. The subfunctionalization is accompanied by prolonged and substantial neofunctionalization in large proportion of duplicate genes. Partial functional relaxation caused by loss of ancestral functions provides the opportunity for advantageous mutations which can lead to new functions.

In summary: the evolution of duplicated genes could go in different directions. The purifying selection, relaxation of purifying selection and positive Darwinian selection affect the evolution of duplicated genes. The action of these forces depends on the original gene function, whether the duplicated genes code for components of multidomain proteins, or if duplicated genes have arisen through a singleton gene duplication or whole genome duplication. In the case of genome duplication duplicated genes are preserved for longer time than in the case of singleton duplication because loss of one of the genes could cause underproduction of its gene product relative to those with which it interacts. As a consequence the genes are preserved for longer time in the genome and have the possibility to evolve a new function.

Subfunctionalization of duplicate genes could offer the possibility for a subsequent evolution of a new function in duplicate genes.

The different evolution and silencing of duplicated genes in geographically separated populations could genetically isolate these populations. Genome duplication produces thousands of duplicated genes (each gene has a duplicate) and a lot of them could be silenced. Therefore, genome duplication could be a source of creation of new species.

The analyses of common carp duplicated *GH* genes, their sequence differences, expression patterns and activity of their products are not only important for a better understanding of growth regulation in this tetraploid fish species but also provide an excellent model to study the evolution of duplicated genes after genome duplication.

The studies and comparison of paralogues of common carp and goldfish are useful for understanding the evolution of duplicated genes after genome duplication. They could answer the question how whole genome duplication could lead to the formation of new species. Because common carp and goldfish most probably arose from the same tetraploid ancestor the studies of these fish species and their subspecies are very interesting. Ohno's

(1970) hypothesis that the increased complexity and genome size of vertebrates has resulted from two rounds (2R) of whole genome duplication in early vertebrate evolution continues to be controversial. The studies of species that experienced a recent whole genome duplication could maybe reject or confirm this hypothesis.

In **Chapter II** of the present study the introns of the second *GH* gene of common carp were sequenced for the first time and a sequence comparison of both *GH* genes was done. It was analysed if both growth hormone genes are polymorphic and a population screening of polymorphisms in *GH* genes was carried out.

In **Chapter III** phylogenetic analyses were done to examine the relationships of common carp *GH* genes with *GH* genes of goldfish and other Cyprinids. In addition, cDNA or gene sequences of common carp and goldfish, some of which also important for growth, were taken from GenBank and phylogenetic analysis was carried out. How duplicated genes have evolved was analysed by applying the relative rate test (Tajima, 1993). Also, it was tested if non-synonymous mutations occur in such a way as to change some amino acid property of interest to a greater extent than is expected under random substitution (neutral evolution) (Hughes et al., 1990).

In **Chapter IV** an online PCR assay was developed to study if there are differences in mRNA levels of the two GH genes of common carp in relation to ontogenesis, sex and season. Samples from 15, 22, 29, 47, 63 and 85 days old fry, one year old fish, three years old males and females and fish adapted to cold water conditions were analysed.

Chapter V describes an elaboration of a new simple method to obtain recombinant active GH proteins without refolding. It allows in future to test whether the activities of products of the two growth hormone genes of common carp are different. Activity was checked by the ability of recombinant GH to stimulate production of IGF - I mRNA in primary common carp hepatocyte cultures.

Chapter II Sequence variation of the two common carp growth hormone genes

II.1 Introduction

Changes of gene expression (gene regulation) after gene duplication appear to be a general rule rather than exception (Wagner, 2000; Gu et al., 2002), and often occur quickly after gene duplication (Gu et al., 2002). These changes appear through mutations in non-coding parts of the paralogues. Initial differentiation of regulatory regions seems to depend more on drift rather than on selection (Lynch & Force, 2000; Zuckerkandl, 2001).

In rainbow trout (another fish species of tetraploid origin) differential levels of the mRNAs for the two GH genes rtGH I and rtGH II in relation to sex and developmental stage (age) have been observed (Yang et al., 1997). Additionally, it has been found that the rtGH II gene does not have an estrogen response element in promoter which is, on the contrary, present in the rtGH I gene.

The effect of polymorphisms in introns for gene regulation was demonstrated by the following example. In pigs a paternally expressed quantitative trait locus (QTL) affecting muscle growth, fat deposition and size of the heart was mapped to the IGF - II (insulin-like growth factor 2) region (Jeon et al., 1999; Nezer et al., 1999). Van Laere (2003) has shown that this QTL is caused by a nucleotide polymorphism in the third intron of IGF - II. The mutation occurs in an evolutionary conserved CpG island that is hypomethylated in skeletal muscle. The mutation abrogates in vitro interaction with a nuclear factor, probably a repressor, and pigs inheriting the mutation from their sire have a threefold increase in IGF - II messenger RNA expression in post-natal muscle. This study establishes a causal relationship between a single-base-pair substitution in a non-coding region and a QTL effect. The result supports the long-held view that regulatory mutations are important for controlling phenotypic variation.

The polymorphism in introns could also affect the creation of secondary structures in precursor mRNA. A mutation occurring in a secondary structural element such as the helix of a RNA hairpin might be individually deleterious because it increases the structure free energy, which may destabilize this structure (Kirby et al., 1995). The secondary structures in pre-mRNA affect splice sites and branch point elements. The presence of structural elements may hinder the accessibility of selected sequences by basic splicing factors. On the other hand, RNA secondary structures that do not involve the conserved splicing sequences

can nonetheless vary the relative distance between these elements. These changes can then determine considerable variation in splice site usage or efficiency (Buratti & Baralle, 2004).

Therefore, differences in promoter and intron regions of the duplicated genes could lead to differential regulation and expression patterns and due to these differences the duplicates could be subfunctionalized (Force et al., 1999; Lynch & Force, 2000). The intron sequences of the GH II gene of common carp as well as promoters of both genes are not sequenced. Therefore, any analysis if differences in promoters and introns of both GH genes could lead to differential regulation of these genes was not possible.

Because of the physiological importance of GH a lot of studies have been performed on the relation between GH gene polymorphisms and quantitative traits in animal species important for man. For example, in cattle relations between some GH gene polymorphisms in the third intron and 3'- end and the quality of milk could be shown (Hoj et al., 1993). Nielsen et al. (1995) demonstrated associations between different alleles in the promoter region of the GH gene and both the regressions of basal plasma GH concentration on age and on weight of pigs.

Grochwska et al. (2001) investigated in cattle genetic variations of plasma GH concentrations and of IGF - I concentration before and following stimulation with thyrotropin -releasing hormone (TRH). A possible association between stimulated GH release, IGF - I and the polymorphism in the GH gene causing a substitution of Leu to Val at amino acid position 127 of the protein was also investigated. The V/V genotypes reached the highest peak value of growth hormone release compared with the other GH genotypes L/L and L/V after TRH injection, whereas L/L genotypes had the highest IGF - I concentrations. It could be possible that the Val genotype is linked with some genotype in a regulator element in promoter or intron that elevates the secretion of the GH gene. Leu genotype could be better in binding with GH receptor. Thereby, Val genotype that could be less favorable for binding with GH receptor could have a compensatory mutation in regulator elements. In humans the associations between GH haplotype and GH expression level have also been demonstrated (Horan et al., 2003).

An impact of GH gene polymorphisms on growth in fish was indicated by a study of Gross & Nilsson (1999) who found significant heterogeneity of the GH I haplotype and genotype frequencies among three size groups of Atlantic salmon (*Salmo salar*) offspring from a hatchery stock. The authors reported that the observed polymorphisms were located in the third intron of the salmon GH I gene.

Wohlfarth et al. (1964) observed differences in growth rates between seven groups of juvenile carp of different origin during winter but similar growth rates during autumn

suggesting that genetic factors and temperature might be able to modify GH secretion. Differences between European and Asian (Chinese) common carp populations have also been reported for growth rate (Moav et al., 1975; Wohlfarth et al., 1975): In poor conditions the fastest relative growth was shown by Chinese carp, while in improved environments European carp showed the better growth.

Since middle of 1990s the genetic variability of carp populations from Europe, Central Asia and East/South-East Asia has been studied extensively using allozymes (Kohlmann et al., 1997; Kohlmann & Kersten, 1999; Kohlmann & Luczynski, 2000; Murakaeva et al., 2003), mtDNA marker genes (Gross et al., 2002) and microsatellites (Kohlmann et al., 2003). All of these studies revealed genetic differentiation among geographical regions, in particular between Europe and East/South-East Asia.

The common carp has a very large natural distribution area (from Central Europe, through Central Asia to East/South-East Asia) with very different ecological conditions (e.g. temperature) and variable growth rates, so that probably genetic varieties of the *GH* gene might be of adaptive importance. In studies such as *GH* gene polymorphisms and quantitative traits in common carp the potential complication caused by the ancestral tetraploid nature of this fish species (it has at least two growth hormone genes) has to be considered. The other problem is that both *GH* genes of common carp are very similar with each other. Therefore, it is difficult to construct specific primer pairs for each of the two growth hormone genes to screen the polymorphisms.

Since the probability to find polymorphisms increases with increasing genetic distance between examined individuals, carp from different geographical regions with known genetic differences were chosen for this part of the study. The tissue collection available at the Leibniz-Institute of Freshwater Ecology and Inland Fisheries (IGB) in Berlin provided the opportunity to study these aspects in wild and domesticated carp from different geographical regions.

II.2 Material and Methods

II.2.1 DNA isolation, PCR, sequencing and cloning conditions

The muscle samples of common carp were taken from the collection of domesticated, feral and wild populations from Europe and Central/East Asia at IGB. For cloning and sequencing four individuals of different geographical origin were selected: one from Europe (EU, wild/feral, River Rhine, Germany), two from Central Asia (CA1 and

CA2, wild, Lake Tuzkan, Uzbekistan) and one from East Asia (EA, domesticated, Wuhan, China).

Total genomic DNA was isolated using the E.Z.N.A. Tissue DNA Kit II (Peqlab Biotechnologie) according to the manufacturer's protocol. Each PCR reaction mix (total 25 μ l) was composed of 1 x PCR buffer (10 mM Tris - HCl, pH 8.3, 50 mM KCl), 2.5 mM $MgCl_2$, 0.1 mM dNTPs, 0.2 μ M of each primer, 5 μ l genomic DNA and 0.25 units of Taq - DNA polymerase (MBI-Fermentas). The Taq - DNA polymerase was used because specific primers for GH II were specific only through one or two nucleotides at the 3' end of the primers. A proofreading polymerase could excise these nucleotides and use GH I DNA as a template.

The sequences of primers and PCR conditions for each primer pair are displayed in Table II.1. The primers were constructed using the Primer-3 software (Rozen & Skaletsky, 2000).

All PCR products were visualized on 1.7% agarose gels using ethidium bromide staining and UV light.

The heteroduplex analysis was carried out according to White et al. (1992).

For sequencing, PCR products were either directly cleaned from the PCR reaction mix using the E.Z.N.A. Cycle-Pure Kit (Peqlab Biotechnologie) or fragments were cut from agarose gels (in case of more than one fragment) and cleaned using the E.Z.N.A. Gel Extraction Kit (Peqlab Biotechnologie) according to the manufacturer's protocols.

For cloning of PCR fragments the TOPO TA cloning kit for Sequencing (Invitrogen) was used. The purification of plasmids was carried out with E.Z.N.A. Plasmid Miniprep Kit I (Peqlab Biotechnologie) or the Qiagen Plasmid Purification Maxi Kit. The automated sequencing was carried out on an ABI Prism 310 Genetic Analyzer (Applied Biosystems) using the DYEnamic ET Terminator Cycle Sequencing Kit (Amersham Pharmacia Biotech) or on a Beckman Coulter CEQ 8000 using the CEQ-DTCS Quick Start Kit (Beckman Coulter). The number of sequenced clones of different PCR products from the four selected samples is presented in Table II.2. The PCR products and fragments inserted in plasmids longer than 700 bp were sequenced using additional internal primers constructed to get overlapping fragments of approx. 700 to 750 bp suitable for automated sequencing. The forward and reverse strands of all fragments were aligned and assembled using SeqEd v.1.0.3. (Applied Biosystems) or Beckman Coulter CEQ 8000 Genetic Analysis System.

Consensus sequences obtained in the present study were submitted to NCBI GenBank. The accession numbers of sequences are given on the page 159.

II.2.2 Comparison of the two common carp GH genes

Sequenced fragments were combined and aligned using BioEdit (Hall, 1999), and differences between sequences were recorded.

The analysis of one amino acid substitution (conservative or not) in one allele of the *GHI* gene was carried out using SCR 3 software (Hughes et al., 1990).

Secondary structure prediction analysis on the basis of multiple alignments of alleles was done with the program RNA secondary structure prediction (Brodsky et al., 1992, 1995) on the GeneBee–Molecular biology server.

The intrinsic DNA curvature in introns was analysed using “DNA Curvature Analysis” software (Gohlke, 1996-2007) on the website <http://www.lfd.uci.edu/~gohlke/curve>.

II.2.3 Population examination on the extent of polymorphism

Genotypes for two length polymorphisms in the third intron of both common carp *GH* genes were deduced from banding patterns of PCR products on agarose gels. Eleven populations with a total number of 278 individuals were screened. Statistical tests for allelic and genotypic differentiation between populations were carried out using GENEPOP (Raymond & Rousset, 1995).

Table II-1 Primer sequences for GH I and/or GH II, length of amplified fragments and PCR conditions.

Primer pair	Gene	Part of the Gene	Length of PCR fragments (bp)	PCR conditions
GH-a: 5'- CCT GAG CGA AAT GGC TAG AG - 3'	GH I	first exon to fifth exon	about 1700, 2100, (both are allele variants)	3 min at 94°C then 35 cycles: 1 min at 94°C, 1 min at 54°C 2 min at 72°C
GH-b: 5'-GCC ATA AGC ACC ACT GTG AA - 3'				
GH-a: 5'- CCT GAG CGA AAT GGC TAG AG - 3'	GH II	first exon to second exon	about 400	3 min at 94°C, then 35 cycles: 30 sec at 94°C, 30 sec at 55°C 1 min at 72°C
GH-k: 5'-CCT GAA CAG GTG TGA CCA ATC - 3' specific primer				
GH-g: 5'-CAG ATA ACC AGC GGC TCT TC - 3'	GH II	second exon to third intron (overlapped with fragment amplified with GH-a + GH-k primer pair)	about 500	3 min at 94°C, then 35 cycles; 30 sec at 94°C 30 sec at 53°C 1 min at 72°C
GH-h: 5'-TGA AAA CCC CTG GAA CAA AG - 3' specific primer				
GH-e: 5'-AAA AAT GAA ATC CCA ACA TGC - 3' specific primer	GH II	third intron to fifth exon (overlapped with fragment amplified with GH-g + GH-h primer pair)	About 1200 or 1500 (both are allele variants)	3 min at 94°C, then 35 cycles; 1 min at 94°C 1 min at 54°C 2 min at 72°C
GH-b: 5'-GCC ATA AGC ACC ACT GTG AA - 3'				
GH-c: 5'-AGG AAC GCA GAC AGC TGA GTAA - 3'	GH I	third to fourth exon	About 430 and 770 (both are allele variants)	3 min at 94°C, then 35 cycles; 30 sec at 94°C 30 sec at 54°C 1 min at 72°C
GH-d: 5'-TAC GGT CAG GCT GTT TGA GA - 3'	GH II		About 650 or 900 (both are allele variants)	

Table II-2 Cloning of PCR products and number of sequenced clones.

Sample	Primer pair	Gene	Approx. fragment length (bp)	Number of sequenced clones
Europe, R. Rhine	GH-a+GH-b	<i>GH I</i>	2100	2
	GH-a+GH-k	<i>GH II</i>	400	2
	GH-g + GH-h	<i>GH II</i>	500	2
	GH-e + GH-b	<i>GH II</i>	1200	2
Central Asia I	GH-a+GH-b	<i>GH I</i>	2100	2
	GH-a+GH-k	<i>GH II</i>	400	2
	GH-g + GH-h	<i>GH II</i>	500	2
	GH-e + GH-b	<i>GH II</i>	1500	2
Central Asia II	GH-a+GH-b	<i>GH I</i>	1700 and 2100	1 1
	GH-a+GH-k	<i>GH II</i>	400	2
	GH-g + GH-h	<i>GH II</i>	500	2
	GH-e + GH-b	<i>GH II</i>	1500	2
East Asia China	GH-a+GH-b	<i>GH I</i>	1700	2
	GH-a+GH-k	<i>GH II</i>	400	2
	GH-g + GH-h	<i>GH II</i>	500	2
	GH-e + GH-b	<i>GH II</i>	1500	2

II.3 Results

II.3.1 Sequencing the introns of GH II and comparison of both GH genes

The coding sequences of the two growth hormone genes of common carp are very similar, therefore it was very difficult to construct primer pairs specific for only one of the two *GH* genes. In order to get intron sequences of the *GH II* gene a primer pair GH-a + GH-b based on the common carp *GH* gene sequences published by Chiou et al. (1990) (*GH I*) and cDNA (Chao et al., 1989) (*GH II*) was constructed (Tab. II.1). The primers lie in conservative regions of both genes. This primer pair was expected to amplify a fragment of

GHI and *GHII* spanning a region from the first to the fifth exon including the whole peptide sequence (Fig. II.1). About 40 individuals were screened after PCR on the agarose gel. Three PCR products were seen: about 1700, 2100 and 2800 bp (Fig. II.2). There were individuals in which only the PCR product of 1700 bp or 2100 bp was obtained and the third variant of all three products was observed in one individual.

The PCR products of 1700 bp and 2100 bp were sequenced successfully. The comparison of their coding regions with two variants of cDNA (Koren et al., 1989; Chao et al., 1989) of common carp and with coding and non-coding regions of the complete *GH* gene sequence (Chiou et al., 1990) showed that they are allelic sequences of the *GHI* gene (cDNA from Koren et al., 1989 and complete gene sequence from Chiou et al., 1990) with different length (large insertion/deletion of 341 bp in third intron) (Appendix: Alignment 1, Tab. A.1). The third PCR product of about 2800 bp was not possible to sequence completely. It was sequenced from the 5' end about 950 bp and from the 3' end about 750 bp. The sequenced parts were identified as *GHI*. The 2800 bp product was found only in the individuals that had both 1700 bp and 2100 bp PCR products (heterozygotes) but not in any one that has only the 1700 bp or 2100 bp PCR product (homozygotes).

In order to clarify the nature of this 2800 bp fragment, artificial heterozygotes were generated by mixing DNA from the respective homozygous individuals before PCR. The resulting PCR products showed a pattern that was identical to heterozygous individuals. Therefore, the additional larger fragment of 2800 bp is a heteroduplex molecule composed of hybridized strands of different sizes, i. e. the short about 1700 bp and long about 2100 bp alleles (Fig. II.1). Figure II.3 represents a schematic image how single strands of short and long alleles could form heteroduplex molecules during renaturation at low temperature

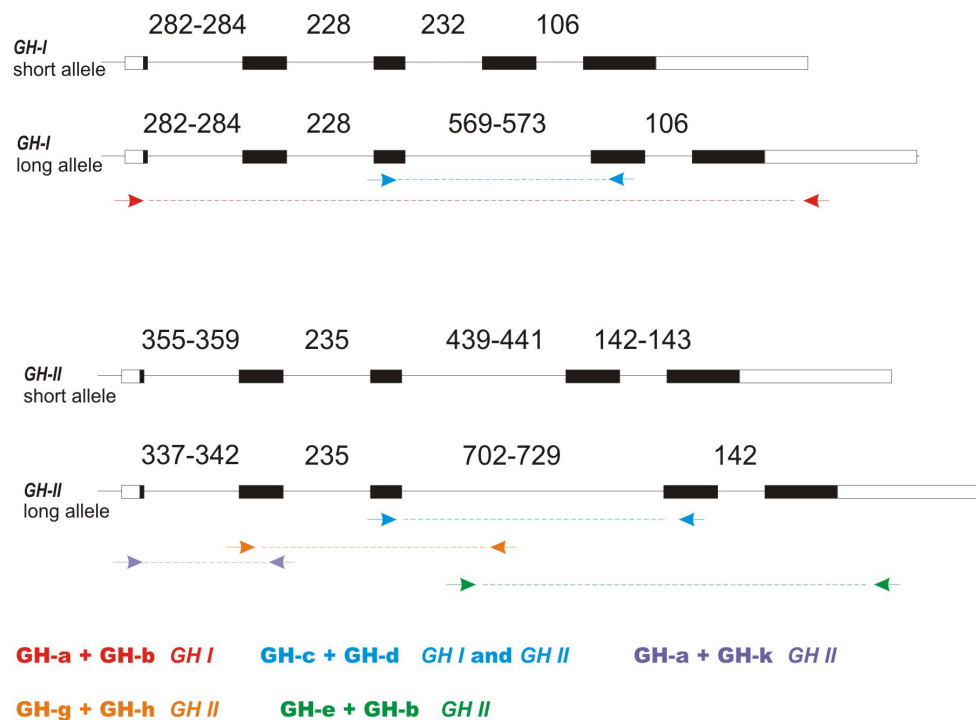


Figure II.1 Primary structure comparison of the two GH genes of common carp. Exons are shown as black boxes (open boxes are untranslated regions) and introns are shown as lines. Numbers are base pairs. The arrows show the primers (colour of arrows according to the colour of primer names)..

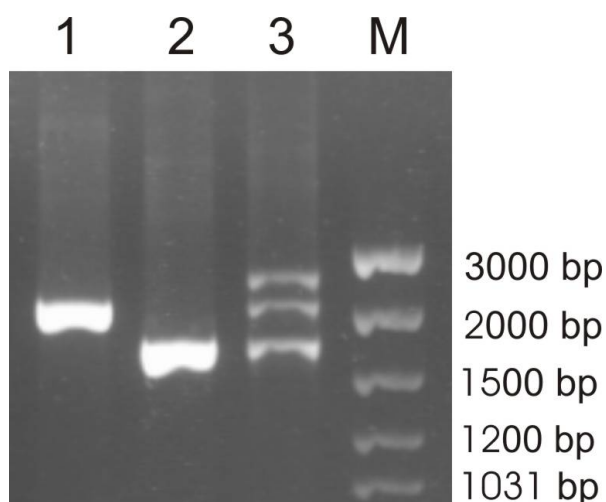


Figure II.2 *GH I* gene fragment patterns of common carp obtained after PCR amplification with primer pair GH-a + GH-b: 1.Homozygote with fragment of about 2100 bp; 2. Homozygote with fragment of about 1700 bp; 3. Heterozygote with fragments of about 1700 bp, 2100 bp and heteroduplex band between them of about 2800 bp. ; M. Molecular weight marker Gene Ruler 100 bp DNA Ladder Plus (MBI-Fermentas)

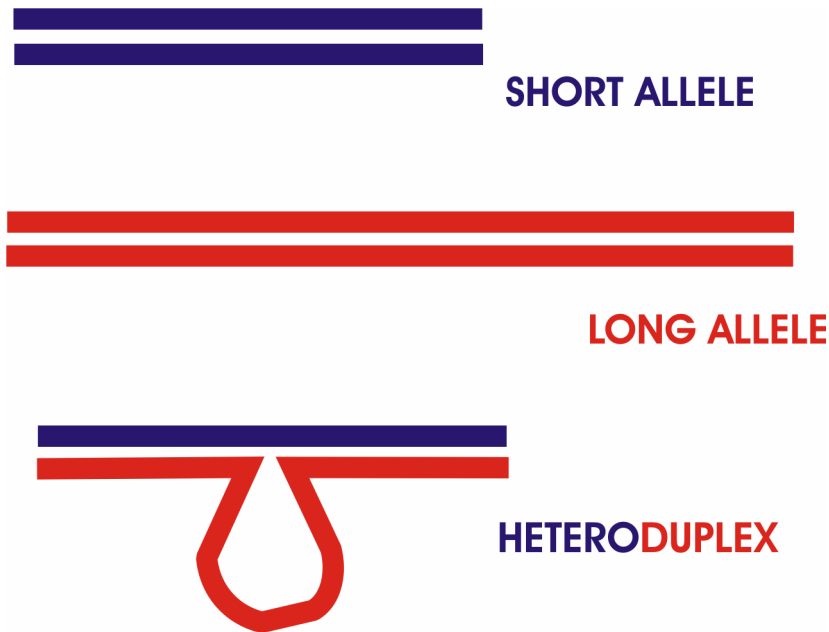


Figure II.3 Heteroduplex scheme. Heteroduplex could appear in heterozygous individuals after denaturation at 90-95°C and then renaturation at lower temperature (usual PCR conditions). Hybridization of long and short single-stranded molecules (from long and short allele variants) forms a loop that could slow down the motion of heteroduplex molecule on the agarose gel.

The *GH II* gene was not amplified at all with primer pair GH-a and GH-b. A second primer pair GH-c + GH-d that should amplify both *GH* genes of common carp was constructed. Both primers lie in very conservative parts of growth hormone genes and span a region from the third exon through third intron to fourth exon (Tab. II.1; Fig. II.1). About 40 individuals were screened on agarose gel after PCR. Three products of different length were obtained: about 430, 770 and 900 bp (Fig. II.4). Two different patterns of PCR products were observed. There were individuals with two PCR products of 770 and 900 bp and a few individuals with three PCR products of 430, 770 and 900 bp. All three PCR products were sequenced successfully. The fragments of 430 and 770 bp were the alleles of the *GH I* gene with the insertion/deletion of 341 bp in third intron. The fragment of 900 bp was the *GH II* gene as shown by comparison with the coding region of both cDNA sequences (Koren et al., 1989; Chao et al., 1989) and the complete gene sequence reported by Chiou et al. (1990).

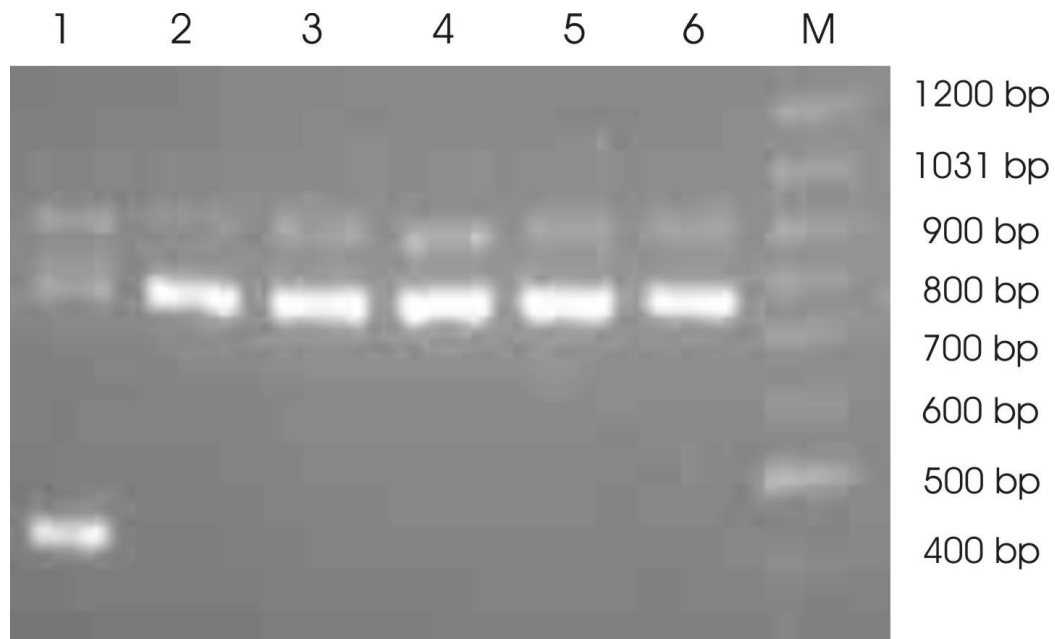


Figure II.4 Common carp GH gene fragment patterns after PCR amplification with primer pair GH-c + GH-d. 1. Individual with three fragments of 430, 770 and 900 bp; 2-6. Individuals with two fragments of 770 and 900 bp; M. Molecular weight marker GeneRuler 100 bp DNA Ladder Plus (MBI-Fermentas)

The primer pair GH-c + GH-d was used in addition for population screening of a length polymorphism (insertion/deletion of 341 bp) at the third intron of *GH I* (see later and Tab II.3). During population screening with this primer pair a very rare fragment of about 650 bp (only two individuals from river Rhine) was found. After sequencing and comparison it was identified as *GH II* with a deletion of 284 bp in the third intron (see later).

The sequence of the third intron of *GH II* obtained with primer pair GH-c + GH-d was different from the sequence of the same intron of *GH I*. The observed differences between the two *GH* genes in the third intron were used to construct specific primer pairs for *GH II*. A lot of primer variants were tested and the following two primers GH-e (specific primer) + GH-b were successful to amplify only the *GH II* gene from beginning of the third intron to 3' untranslated region (Tab. II.1). This primer pair produced fragments of about 1200 or 1500 bp in different individuals (Fig. II.5). The fragment of 1500 bp was frequent and that of 1200 bp was rare: only in two individuals with a 284 bp deletion in third intron of the *GH II* gene.

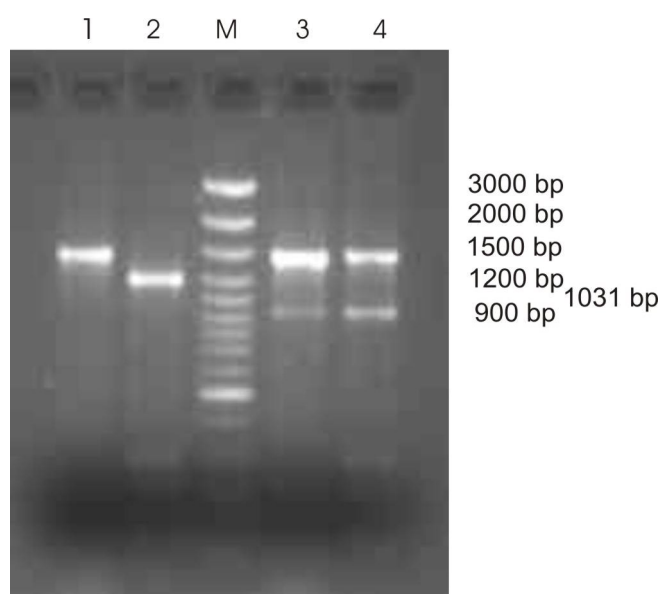


Figure II.5 Common carp GH II gene fragment patterns after PCR amplification with primer pair GH-e + GH-b. 1. Homozygote for fragment of about 1500 bp; 2. Homozygote for fragment of about 1200 bp; M. Molecular weight marker GeneRuler 100 bp DNA Ladder Plus (MBI-Fermentas); 3,4. Homozygotes for fragment of about 1500 bp. The short fragments of about 950 bp in individuals 3 and 4 were also sequenced, but they did not have any similarity with *GH* genes

It was not possible to construct a single specific primer pair for GH II from the first exon to the third intron. A lot of primer variants were also tested. Two primer pairs that were specific for shorter regions of the GH II gene were constructed. One primer pair was GH-g + GH-h (specific primer) (Tab. II.1). The fragment obtained with this primer pair overlapped with the fragment obtained with GH-e and GH-b and amplifies the part of GH II from the second exon to the third intron (Fig. II.1). The length of this fragment was about 500 bp.

The last primer pair GH-a + GH-k (specific primer) specific for GH II amplified the fragment from the first exon to the second exon and overlapped with the fragment obtained with the GH-g + GH-h primer pair. The size of this fragment was 400 bp.

Figure II.6 shows a schematic representation of four PCR patterns obtained with primer pair GH-c + GH-d. This primer pair amplifies both growth hormone genes from the third exon through the third intron to the fourth exon. For each gene two alleles with insertion/deletion in the third intron were found. All four PCR products had different lengths and could be well separated from each other on the agarose gel. On the scheme alleles of each gene are represented by different colour. The first fragment pattern was found only in two fish from river Rhine. The second and third patterns were found in European as well as

in Asian populations. The second pattern was more frequent in Europe and Central Asia and the third in the East Asian population from China. The fourth fragment pattern was found only in the Chinese population.

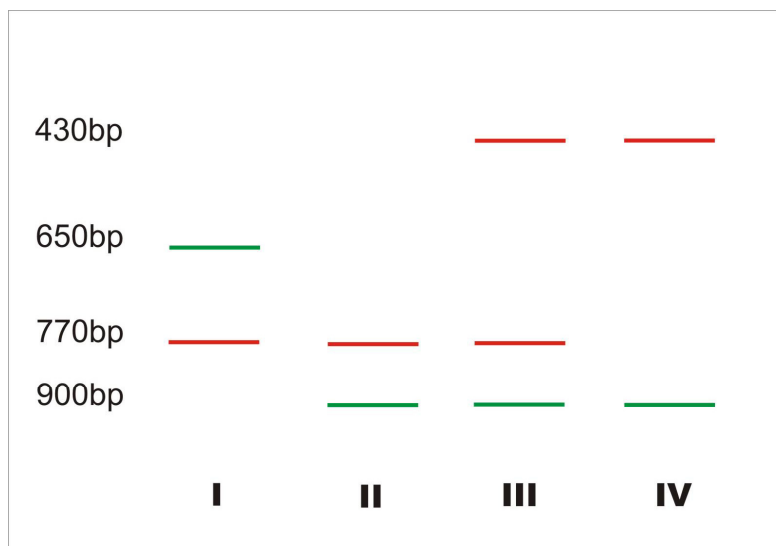


Figure II.6 Fragment patterns obtained after PCR amplification with primer pair GH-c + GH-d. Red bands are alleles of GH I and green bands are alleles of GH II. I - fish from river Rhine; II - first fish from Central Asia I; III - second fish from Central Asia II; and IV - fish from East Asia (China).

Four carp samples of different origin (Europe and Central/East Asia) that showed length polymorphisms in the third intron of the GH I and GH II genes were selected (fish from river Rhine - first pattern, fish from Central Asia I - second pattern, fish from Central Asia II - third pattern and fish from East Asia (China) - fourth pattern) to obtain complete GH gene sequences. The PCR was done with DNA from all four samples using the following primer pairs: GH-a + GH-b (GH-I); and GH-e + GH-b, GH-g + GH-h, GH-a + GH-k (GH-II). The PCR products were sequenced. At the same time some amount of PCR products were cloned and also sequenced (Tab. II.2). The results of sequencing of PCR products and of plasmids were compared with each other to avoid PCR errors. The consensus sequences for GH I and GH II are presented in Alignments 1 and 2 of Appendix. To better represent the differences between alleles of the two growth hormone genes, they were entered in Table A.1 for GH I alleles and in Table A.2 for GH II alleles of Appendix.

The comparison of cDNA (Koren et al., 1989 (*GH-I*); Chao et al., 1989 (*GH-II*)) and complete gene sequence (Chiou et al., 1990 (*GH-I*)) with consensus sequences for *GH I* and *GH II* obtained in the present studies was done. Only in one allele of the *GH I* gene in fish

from Central Asia (Central Asia I in the long allele) one amino acid substitution from phenylalanine to isoleucine Phe/Ile at position 84 of mature protein was found (Appendix: Alignment 3; Tab. A.1). However, analysis with SCR3 has shown that this amino acid substitution is conservative. In all allelic variants of *GH II* obtained in the present study no amino acid substitutions were found and the mature protein was identical with mature protein from the cDNA from Chao et al. (1989). All obtained *GH II* gene fragments from the beginning of the third intron to the 3' untranslated region were identical in the 3' non-coding end at nucleotide positions from 2159 to 2163 but demonstrated differences to the cDNA reported by Chao et al. (1989) (Alignment 2 in Appendix). It could be possible that the Chao et al. (1989) cDNA sequence has sequencing mistakes in 3' end maybe due to imperfection of sequencing techniques used at that time. Silent substitutions were found in alleles of both duplicated genes (Appendix: Tab. A.1 and Tab. A.2).

All the introns in the *GH II* gene similar as in the *GH I* gene of common carp start with a consensus GT dinucleotide and end with AG. The arrangement of exons and introns in the *GH II* gene is the same as in the *GH I* gene – five exons and four introns (Fig. II.1). The introns of the *GH II* gene were longer than in the *GH I* gene (Fig. II.1): The first intron varied in the *GH I* gene from 282 to 284 bp and in the *GH II* gene from 337 to 359 bp; the second intron in *GH I* gene was 228 bp and in *GH II* gene 235 bp long; the third intron in *GH I* gene varied from 232 to 573 bp and in *GH II* gene from 439 to 729 bp, the fourth intron in *GH I* gene was 106 bp and in *GH II* gene 142-143 bp long.

The A+T content is higher than G+C content in all introns of both *GH* genes of common carp. The similarity of the introns is 59.5-70.1% for first intron, 82% for second intron, 22.5-61.9% for third intron and 57.5% for fourth intron. In coding sequence similarity is 96%.

In the first intron of the *GH II* gene four highly variable sites with microsatellites were found (Tab. A.2 and Alignment 2 in Appendix):

1. (CA)₁₃, (CA)₈ or CA
2. (TA)₃₂T(TA)₁, (TA)₃₂A, (TA)₆TG(TA)₅TC(TA)₆GATCA,
(TA)₁₄TC(TA)₆GATCA, (TA)₁₉TCA, (TA)₂₀TCA or (TA)₁₈TCA
3. (TA)₅T₆CT₇, (TA)₅T₇CT₆, (TA)₃T₁₁CT₇ or (TA)₅T₈CT₇
4. (TTA)₇T₁₀, (TTA)₄T₁₇, (TTA)₃T₁₂, (TTA)₃T₁₃, or (TTA)₅T₁₂,

In the corresponding intron of the *GH I* gene only a very short microsatellite motif (TA)₂ and poly A repeat were found (Tab. A.1 and Alignment 1 in Appendix).

The secondary structure prediction in the first intron of the *GH II* gene on the basis of multiple alignments of alleles has shown that the microsatellite motif of the second site could form a hairpin. The model energy of structure was from -12.8 Kkal/mol to -30.6 Kkal/mol (Fig. II.7). The model energy includes inputs from conservative and complementary pairs with corresponding coefficients.

There was one deletion of 49 bp in the first intron of the *GH II* gene in fish from East Asian origin that was not found in any allele from fish of European and Central Asian origin. In the first intron of the *GH I* gene no deletion was found.

The preliminary analysis of introns from both duplicated genes with computer software has demonstrated that in some of them DNA curvature could be found. The DNA curvature in the first intron of *GH I* is remarkably stronger than in the same intron of *GH II* (Appendix: Fig. A.1; A.2; A.3). In the latter it is feeble. The second and fourth introns of both genes also demonstrate feeble DNA curvature. The large deletion of 341 bp in the third intron of the *GH I* gene reduces DNA curvature in this intron (Appendix: Fig. A.4; A.5). On the other hand, the large deletion of 284 bp in the same intron of *GH II* does not reduce DNA curvature (Appendix: Fig. A.6; A.7).

(Europe, Central Asia or East Asia) and status (domesticated, wild/feral, wild). All of the five European domesticated populations originating from Germany, Poland, Czech Republic, Hungary and one wild/feral population from Germany (River Rhine) were found to be fixed for the long allele (Tab. II.3). The remaining German wild/feral population (River Danube) as well as the two Central Asian wild populations from Uzbekistan and the captive stock of Russian R. Amur wild carp displayed polymorphism with the long allele clearly dominating (frequencies higher than 0.933). The domesticated population from Wuhan, China, showed a lower frequency of the long allele (0.737) and was the only population in which a homozygote for the short allele could be detected. Statistical tests for allelic and genotypic differentiation between populations revealed that the Chinese population differed significantly from all other populations whereas the differences between all other populations were not significant. All studied polymorphic populations were found to be in Hardy-Weinberg equilibrium.

In all of the same populations that were examined for this *GHI* polymorphism only two fish with a deletion of 284 bp in the third intron of the *GHI* gene were found, but as homozygotes. These two fish were from the R. Rhine (Germany) wild/feral population.

Table II-3 Genotype and allele frequencies at the *GHI* gene in common carp populations of different geographical origin and status (L = long allele; S = short allele).

Population	Sample size	Frequencies				
		Genotypes			Alleles	
		LL	LS	SS	L	S
Europe						
R. Rhine (Germany, wild/feral)	27	1.00			1.00	
R. Danube (Germany, wild/feral)	30	0.867	0.133		0.933	0.067
Petershain (Germany, domesticated)	27	1.000			1.000	
Fiedler (Germany, domesticated)	18	1.000			1.000	
Zator (Poland, domesticated)	28	1.000			1.000	
Pohorelice (Czech Republic, domesticated)	28	1.000			1.000	
Tata (Hungary, domesticated)	16	1.000			1.000	
Central Asia						
L. Tuzkan (Uzbekistan, wild)	29	0.966	0.034		0.983	0.017
L. Aidar (Uzbekistan, wild)	30	0.967	0.033		0.983	0.017
East Asia						
R. Amur (captive stock of a Russian wild population)	26	0.923	0.077		0.962	0.038
Wuhan (China, domesticated)	19	0.526	0.421	0.053	0.737	0.263

II.4 Discussion

The *GH II* gene of common carp could not be amplified with primer pair GH-a + GH-b, though the primers were specific to both growth hormone genes. The explanation for this could be: the *GH II* gene fragment that should be amplified with this primer pair is some hundred base pairs longer than the *GH I* gene fragment. All introns of the *GH II* gene are longer than those of the *GH I* gene (Fig. II.1). Shorter fragments have privilege in PCR in comparison to longer fragments. This might be due to competition for primers and nucleotides between fragments. Also, the shorter fragments could be amplified faster leading to short allele dominance and large allele dropout (Wattier et al., 1998). On the other hand, there is a microsatellite motif in the first intron of the *GH II* gene that could form a secondary structure which might hinder the amplification of long fragments from the *GH II* gene. The secondary structure formation could also explain why it was not possible to obtain the whole first part of the *GH II* gene from the first exon to the third intron.

The arrangement of exons and introns in the *GH II* gene is the same as in the *GH I* gene. On this basis, the possibility that growth hormone gene duplication in common carp was caused by a retrotransposition of mRNA could be ruled out.

The low similarity of *GH I* and *GH II* gene introns shows that in introns the substitution rate is higher than within a coding part. The 3' and 5' ends of all four introns are better conserved between the two paralogues than the middle part. This could be explained by the important role that 3' and 5' ends of introns play for the splicing process (Berget, 1995; Berglund et al., 1997; Hertel et al., 1997; Moore, 2000; Smith & Valcarcel, 2000).

Only in one allele of *GH I* from one Central Asian fish an amino acid substitution was found. But it was conservative. In contrast, none of the sequenced alleles from *GH II* showed any amino acid substitution. It could be possible that mutations are deleterious even if they occur in only one of paralogues. The purifying selection could act against changes at critical sites in either copy.

It has been demonstrated that the first intron of several genes such as $\alpha 1(I)$ Collagen of human, human platelet derived growth factor-B, human Interferon- γ and human epidermal growth factor receptor has an important regulatory function (Bornstein et al., 1988; Franklin et al., 1991; Sica et al., 1992; Chrysogelos, 1993).

The differences between the two growth hormone genes of common carp in first intron are very large. There are four different microsatellite sites in first intron of *GH II*. They were not found or were very short in first intron of *GH I*. The microsatellites are most

generally considered in terms of their roles as genetic markers for studies in population genetics, evolutionary relationships, and gene mapping. However, there is much evidence accumulating that microsatellite sequences also serve functional roles as coding or regulatory elements (Kunzler et al., 1995; Kashi et al., 1997). A tetranucleotide polymorphic microsatellite (TCAT)_n located in the first intron of the tyrosine hydroxylase gene in human acts as a transcription regulatory element in vitro. In electrophoretic mobility shift assays these tetrameric repeated sequences form specific complexes with HeLa cell nuclear extracts (Meloni, 1998). Gebhardt et al. (1999) investigated modulation of epidermal growth factor receptor (EGFR) gene transcription by a polymorphic dinucleotide repeat CA in intron I. There were alleles with 16, 17, 18, 20 and 21 CA repeats. The authors could demonstrate that transcription in vitro and in vivo of the EGFR gene is inhibited in alleles with 21 CA repeats.

A dinucleotide repeat CA was also found in the first intron of common carp GH II but was absent in GH I. There were one repeat in fish from Central Asia, 8 repeats in fish from Europe (Rhine) and 13 repeats in fish from East Asia (China) (Appendix: Alignment A.2, Tab. A.2). It should be compared if differences in presence of CA microsatellites in first introns of two growth hormone genes could lead to differences in expression of these genes. Furthermore it should be tested if the polymorphism of this dinucleotide repeat could affect the regulation of the GH II gene. Possibly, some QTLs in common carp could be associated with these dinucleotide repeats in first intron of GH II. The polymorphism in a multiallelic microsatellite marker CA located at the 5' flanking region of the Insulin-like growth factor-1 gene was associated with growth and carcass traits in swine. In the preliminary analysis, significant associations were observed (Casas-Carrillo et al., 1997).

In the first intron of GH I only a very short repeat (TA)₂ is present. On the contrary, in the first intron of GH II (TA) was repeated from 6 to 32 times in different alleles (Alignments A.1 and A.2; Tab. A.1 and A.2 of Appendix). On the basis of multiple alignments of the GH II gene alleles a hairpin motif formed by microsatellites TA in first intron was predicted. In introns secondary structures could play an important role for splicing or affect the accessibility of selected sequences by basic splicing factors (Kirby et al., 1995; Buratti & Baralle, 2002). Therefore, due to differences of the microsatellite composition in the first intron the processing of pre-mRNA of the two growth hormone genes could proceed with different intensity. That could finally lead to different ratios of mRNA for GH I and GH II.

The ability of DNA to bend in sequence-dependent manner has an important function in many biological processes like DNA replication, site-specific recombination,

and transcription (Travers & Klug, 1990). Helical conformation analysis (Gabrielian & Pongor, 1996; Gabrielian et al., 1996) on the basis of the (CA)_n and (TA)_n repeats and flanking sequences indicate that these dinucleotide repeats are highly flexible. The intrinsic DNA curvature propensity, a measure for helical asymmetry frequently associated with a rigid conformation, is remarkably low for poly-CA and poly-TA stretch. But the authors have found that bendability is prominently elevated. The longer the poly-CA or poly-TA stretch, the longer the highly bendable section becomes, too. This could also favour a DNA secondary structure that supports or hinders binding of a transcription regulation factor. In addition, the bendability of (TA)_n repeats is in fact well known from experiment (Calladine & Drew, 1992). Also, protein-induced bending at TATA sequences seems to be especially pronounced (Simpson & Kunzler, 1979; Kim et al., 1993; Kim & Burley, 1994; Shakked et al., 1994; Starr et al., 1995).

Viewed as functional elements of the genome, the special characteristics of microsatellites as mutational hotspots have led to the proposal that microsatellites may be a major source of quantitative genetic variation and evolutionary adaptation (Kashi et al., 1990, 1997; King, 1994). This could enable a population to replenish genetic variation lost through drift or selection, and serving as adjustable tuning knobs through which specific genes are able to rapidly adjust the norm of reaction to minor or major shifts in evolutionary demands (King et al., 1997; King & Soller, 1999). In case of diversification of duplicated genes the microsatellites could lead to differences in gene transcription and processing of pre-mRNA of duplicates. Because of this, the paralogous GH genes of common carp could have different expression patterns.

The other feature of the DNA molecule, the sequence-dependent DNA curvature, is known to play an important role in the initiation of transcription of many genes (Trifonov, 1985; Hagerman, 1990; Harrington, 1992; Perez-Martin et al., 1994). The preliminary analysis that was done in the present study with computer software found curvature DNA in non-coding parts of both growth hormone genes. The higher intensity of DNA curvature found in first intron of GH I in comparison with the first intron of GH II gene could be explained by sequence differences between the two introns. Curved DNA molecules result if special base sequences or structural motifs are repeated in phase with the DNA helical repeat. The special effect is produced by runs of homopolymeric dA.dT base pairs ("A-tracts"). The poly-A and poly-T are in fact known to be rigid (Rhodes, 1979). Periodical combination of short rigid tracts spaced apart with short flexible regions in the DNA molecule lead to DNA curvature (Gabrielian et al., 1996). The curved DNA was discovered during the study of a minicircle DNA from kinetoplast body of *Leishmania tarentolae*, in

which appropriately phased A-tracts or T-tracts occur and are spaced apart (Marini et al., 1982; Diekmann & Wang, 1985; Crothers et al., 1990). In the first intron of GH I a pattern like this could be seen. The first intron of GH II has relatively long microsatellites of poly-CA and poly-TA that are very flexible as mentioned above. This could prevent DNA curvature because it interrupts functional motifs that are repeated in phase with the DNA helical repeat.

The large deletion of 341 bp in third intron of the GH I gene impacts the intrinsic DNA curvature of this intron. With the 341 bp deletion it is less curved than without the deletion. Possibly, this deletion could affect the expression of GH I. Suzuki & Yagi (1995) have demonstrated that even a small intrinsic curvature may greatly enhance the affinity of protein-DNA contacts. Matthews (1992) has suggested that DNA curvature or looping bring together components of the transcriptional complex that are distant along the DNA sequence. Thus, possibly the differences in DNA curvature in introns between paralogues as well as between alleles of the same gene could lead to differences in gene regulation.

The population examination on the extent of polymorphism at the large insertion/deletion (341 bp) in third intron of the *GHI* gene has shown that carp populations from Europe and Asia differ significantly in the frequencies of the short allele. All five domesticated populations from Europe were fixed for the long allele. The domesticated population from China showed the highest frequency of the short allele and was the only one population in which a homozygote for the short allele could be detected. The populations were found to be in Hardy-Weinberg equilibrium and random genetic drift could not be excluded. But it was demonstrated by Moav et al. (1975) and Wohlfarth et al. (1975) that in poor conditions the fastest relative growth was shown by Chinese carp and in improved environments European carp showed the better growth. Therefore, it could be possible that differences in allele frequencies between European and Asian populations lead to differences in relative growth under different environmental conditions in populations from Europe and China. Another possible explanation is that European domesticated carp were selected for fast growth rate (mainly by mass selection) – the long allele could thus be fixed if it is associated with fast growth. As it was mentioned above, the short and long allele of *GHI* could have an impact on the regulation of this gene because they show different intensity of DNA curvature in the third intron.

The large deletion of 284 bp in the third intron of the GH II gene was found in only one population – the wild/feral European population from Rhine in only two fish. In the present study it was a very rare allele. Surprisingly, both fish were homozygotes for this allele of GH II. This could also be explained with short allele dominance and large allele

dropout in PCR (Wattier et al., 1998). In future more wild populations from Europe should be studied to examine how frequent the short allele is.

In the future it would be useful to investigate the effect of microsatellite patterns or number of repeats in the first intron of the GH II gene on transcription or pre-mRNA processing of this gene and if it could lead to different expression patterns of duplicated growth hormone genes. An analysis of transcription activity in vitro should be done.

An experimental analysis should be carried out if intrinsic DNA curvature is present in introns of duplicated GH genes. If this will be confirmed, it would be interesting to examine whether differences in DNA curvature between the first introns of both genes and if the large deletion of 341 bp in the third intron of GH I that decrease DNA curvature could affect gene regulation.

Also, it should be tested if polymorphisms of GH I and GH II could be related to some quantitative traits in common carp, in particular to growth rate.

The promoters of both growth hormone genes of common carp were not sequenced in this study. It should also be done in future to see if the genes also have differences in the promoter region that could affect the expression of these duplicate genes.

Chapter III Phylogeny and evolution of duplicated genes of common carp and goldfish

III.1 Introduction

The phylogenetic approach can be used for the analysis of relationships between the two growth hormone genes of common carp. The lineage specific genome duplication in common carp predicts that common carp should have two orthologues for each gene of other diploid cyprinid species like grass carp, *Ctenopharyngodon idella*; bighead carp, *Hypophthalmichthys nobilis*; silver carp, *Hypophthalmichthys molitrix*; zebrafish, *Danio rerio* and diploid species from other animal classes. Since common carp and goldfish have the same tetraploid ancestor the phylogenetic analysis of growth hormone sequences of these two species is very useful for understanding the nature of duplicated growth hormone genes in these species. The goldfish should also have two orthologues for each gene of other diploid species. If growth hormone gene duplication has already occurred in ancestor of common carp and goldfish then one of the *GH* duplicates of common carp should be a sister sequence to one of the goldfish duplicates (they should be orthologues) and the second to the other duplicate of goldfish (second orthologues of *GH* genes of common carp and goldfish). The cDNA sequences for two variants of growth hormone of goldfish are published in GenBank (accession numbers AF401272 and AF401273).

After whole genome duplication other genes of common carp and goldfish should show the same gene tree topology that was described above. Therefore, it should be searched for other duplicated common carp and goldfish genes in GenBank and phylogenetic trees should be constructed to compare them with each other.

The phylogenetic approach can also help to see if other genes that play an important role in growth and development have more than one copy in common carp and goldfish.

In addition, it can be analysed how duplicated genes evolved. The mutations in coding regions could be of two types: synonymous (which do not change the amino acid) or non-synonymous (which do change the amino acid). Over evolutionary time, changes of these two types occur at very different rates, and comparing these rates can be used to provide information regarding the action of natural selection (Nei, 1987). A comparison of synonymous and non-synonymous substitutions in duplicated genes of varying ages and from a diversity of species suggests that genes experience a period of accelerated evolution shortly after gene duplication (Lynch & Conery, 2000). Goodman (1981a,b) and Czelusniak et al. (1982) contended that the rate of evolution often accelerates following gene

duplication. For example, they claimed that extremely high rates of amino acid replacements occurred following the gene duplication that gave rise to the α - and β -hemoglobins, and that the high rates were due to advantageous mutations that improved the function of these globulin chains.

For studies of natural selection at the molecular level, it is useful to obtain separate estimates of the number of synonymous nucleotide substitutions per synonymous site (designated dS) and of the number of non-synonymous nucleotide substitutions per non-synonymous site (designated dN) (Nei & Gojobori, 1986).

Since most amino acid changes are disadvantageous, synonymous substitutions occur at a higher rate than non-synonymous ones in most genes, due to purifying selection and the ratio dN : dS should be <1 . Under neutral evolution, the rates of synonymous and non-synonymous substitutions are expected to be equal and the ratio dN : dS should be $=1$ (Kimura, 1983).

Positive Darwinian selection can be responsible for functional divergence between the duplicated genes (e.g. Zhang et al., 1998; Duda & Palumbi, 1999; Hughes et al., 2000). Under positive Darwinian selection, natural selection favours amino acid replacements. As a result, non-synonymous mutations get fixed at a faster rate than synonymous mutations (ratio dN : dS >1), as has been shown for genes such as primate lysozyme genes (Messier & Stewart, 1997), pregnancy-associated glycoprotein genes (Hughes et al., 2000), primate ribonuclease genes (Zhang et al., 1998), and conotoxin genes (Duda & Palumbi, 1999).

Neofunctionalization of duplicated genes requires varying numbers of amino acid substitutions. The functional change in primate ribonuclease genes probably required many substitutions (Zhang et al., 1998), but the functional difference between the red and green opsins is largely attributable to only two substitutions (Asenjo et al., 1994). Therefore, if for the change of gene function only a few amino acid substitutions are needed then the ratio dN : dS could not be >1 , although positive Darwinian selection favours these few amino acid substitutions. On the other hand, once the gene has adapted to its specific function, purifying selection is expected to predominate, allowing the number of synonymous substitutions per site to catch up and eventually exceed the number of non-synonymous substitutions per site (Hughes, 1999; Hughes et al., 2000).

A relative rate test (Tajima, 1993) could also be applied to examine if one of the duplicates evolved at a faster rate after duplication, which could point to either that functional constraints have been relaxed at one point in time during functional divergence or that positive Darwinian selection has occurred.

In addition, Hughes et al. (1990) developed an alternative method for testing whether sequences have been subjected to positive Darwinian selection by evaluating whether non-synonymous mutations occur in such a way as to change protein hydrophobicity, charge or polarity to a greater extent than is expected under random substitution. This method involves the computation of the proportion of radical non-synonymous differences per radical non-synonymous site (pNR) versus the proportion of conservative non-synonymous differences per conservative non-synonymous site (pNC). If $pNR > pNC$ then non-synonymous differences occur in such a way as to change the property of interest to a greater extent than expected at random. Such a nonrandom pattern of amino acid replacement is suggestive of natural selection favouring changes in the residue property.

III.2 Material and Methods

The gene sequences of common carp, goldfish, grass carp, silver carp, bighead carp, wuchang bream, black Amur bream, zebrafish, chicken (*Gallus gallus*), domestic cow (*Bos taurus*), human et cetera were obtained by BLASTN from GenBank. The common carp sequences were used as BLASTN query sequences using sets: Choose Search Set – others (nc etc); Program Selection – more dissimilar sequences (discontiguous megablast). Distance tree of results option in BLAST results was used to see preliminary phylogenetic trees. Accession numbers of the sequences used for dendrogram construction in the present study can be found on Fig. III.1 - Fig. III.9 (dendrograms). The chromosome numbers in species used for phylogenetic analysis could be found in Table III.1.

Sequences were aligned by MEGA 3 program (Kumar et al., 2004) using ClustalW algorithm (Thompson et al., 1994). Editing of the alignments was done with the BioEdit sequence editor. The nucleic acid sequences were aligned using the amino acid alignments as guides. If the alignment postulated a gap in any sequence, the corresponding codons were removed from all sequences used for dendrogram construction. For phylogenetic analysis only coding parts of sequences were used.

Phylogenetic trees were reconstructed by the MEGA 3 program using the Neighbor-Joining method (Saitou & Nei, 1987). Evolutionary distances were computed according to Kimura (1983), Tajima-Nei (1984), Tamura-Nei (1993) and Tamura-Nei γ (1993) methods. Bootstrap values (1000 replications) were obtained according to Felsenstein (1985). The Phylo-Win program (Galtier et al., 1996) was used to confirm the dendrogram topology using Maximum Likelihood method. Bootstrap values were also based on 1000 replications.

Numbers of synonymous nucleotide substitutions per synonymous site (d_S) and non-synonymous substitutions per non-synonymous site (d_N) were estimated by Nei &

Gojobori's (1986) methods using one parameter model of Jukes-Cantor (1969) implemented in MEGA 3. For this analysis pairwise alignments were done.

The divergence time between paralogues of common carp and paralogues of goldfish was calculated using the formula $T=K/2r$, where T is time of divergence, K is the number of synonymous substitutions per synonymous site between two analysed homologous sequences, and r is the rate of substitution. Substitution rates (r) for silent (synonymous) sites were taken from Li & Graur (1991): 4.61×10^{-9} as an average for 26 genes. These rates are based upon comparisons between human and mouse/rat genes, assuming that the time of divergence was 80 Myr ago (Li & Graur, 1991). Number of synonymous substitutions per synonymous sites is used for divergence time definition since these substitutions do not result in amino acid replacements and therefore they should be neutral.

Relative rate test was applied using the nonparametric rate test (Nei & Kumar, 2000) developed by Tajima (1993) and implemented in MEGA 3. The advantage of using a nonparametric test is that the results are not influenced by the choice of a possibly wrong substitution model (Nei & Kumar, 2000). The relative rate test of Tajima compares two sequences with an outgroup sequence and counts the number of unique substitutions in both lineages. If both genes evolve under the molecular clock hypothesis, both genes are expected to have accumulated a similar number of 'unique' substitutions (Tajima, 1993; Nei & Kumar, 2000). On the other hand, if one of the duplicates has accumulated a significantly larger number of substitutions, the molecular clock does not apply and one of the paralogues is inferred to have experienced an increased evolutionary rate. To compare relative evolution rates between common carp or between goldfish paralogues the appropriate zebrafish orthologue was taken. Also, the zebrafish orthologue was used to compare relative evolution rates between some orthologous genes from common carp and goldfish. The applied relative test was based both on the amino acid sequences and on the corresponding nucleic acid sequences using first, second and third codon positions together and separately.

In addition, a test of positive Darwinian selection was applied to duplicated genes. It was tested if non-synonymous mutations occur in such a way as to change some amino acid property of interest to a greater extent than is expected under random substitution (neutral evolution). For this approach the program SCR3 (Hughes et al., 1990) was used.

III.3 Results

The coding sequences of common carp and goldfish *GH* cDNA were used for a phylogenetic analysis with *GH* coding sequences from other diploid cyprinid species. The number of chromosomes in fish species used for dendrogram construction could be found in

Table III.1. The coding sequences of common carp paralogues used for phylogenetic analysis were taken from Koren et al. (1989) (*GH I*) and Chao et al. (1989) (*GH II*). In Chapter II of the present study additional alleles for *GH I* and *GH II* genes of common carp have been described. However, all but one alleles of *GH I* coded a protein identical with cDNA from Koren et al. (1989). The exception was an allele with one conservative amino acid substitution from phenylalanine to isoleucine Phe/Ile at position 84 of mature protein. All alleles of *GH II* coded a protein identical with the cDNA sequence published by Chao et al. (1989).

The phylogenetic analysis of GH genes from cyprinid fishes included two genes from common carp and goldfish, one gene of zebrafish *Danio rerio*, grass carp *Ctenopharyngodon idella*, silver carp *Hypophthalmichthys molitrix*, bighead carp *Hypophthalmichthys nobilis*, wuchang bream *Megalobrama amblycephala* and black Amur bream *Megalobrama terminalis*. The Neighbor-Joining method using the Kimura two-parameter distance produced a phylogenetic tree with high bootstrap support (Fig. III.1). It displays two common carp and two goldfish genes belonging to one monophyletic group of 99% bootstrap support. In addition, each of two GH genes of common carp is an orthologous sequence to one of two GH genes of goldfish. The orthologues of common carp and goldfish GH genes group together with 90% (GH I) and 98% (GH II) of bootstrap support, respectively.

Virtually identical trees are obtained if for the Neighbor-Joining method Tajima-Nei, Tamura-Nei or Tamura-Nei- γ distances are used. The same tree topology was obtained with the Maximum Likelihood method using Kimura two-parameter distance.

Table III-1 Chromosome numbers in species used in this study for phylogenetic analysis

Species	Klass or family	Diploid/zygotic(2n)	Reference
Cyprinus carpio	Cyprinids	98-104	Arkhipchuk, 1999; Klinkhardt et al., 1995
Carassius auratus		100	Klinkhardt et al., 1995
Danio rerio		50	Arkhipchuk, 1999; Klinkhardt et al., 1995
Megalobrama amblycephala		48	Arkhipchuk, 1999; Klinkhardt et al., 1995
Megalobrama terminalis		48	Arkhipchuk, 1999
Ctenopharyngodon idella		48	Arkhipchuk, 1999; Klinkhardt et al., 1995
Hypophthalmichthys molitrix		48	Arkhipchuk, 1999; Klinkhardt et al., 1995
Hypophthalmichthys nobilis (Aristichthys nobilis)		48	Arkhipchuk, 1999; Klinkhardt et al., 1995
Pimephales promelas		50	Arkhipchuk, 1999; Klinkhardt et al., 1995
Psetta maxima		40-44	Arkhipchuk, 1999; Klinkhardt et al., 1995
Epinephelus coioides	Scophthalmids (Turbots) Serranidae	no study	
Typhlonectes natans	Amphibia	no study	
Gallus gallus	Avis	78	NCBI Chicken Genome Resource
Bos taurus	Mammalia	60	NCBI Bovine Genome Resource
Homo sapiens		46	NCBI Human Genome Resource

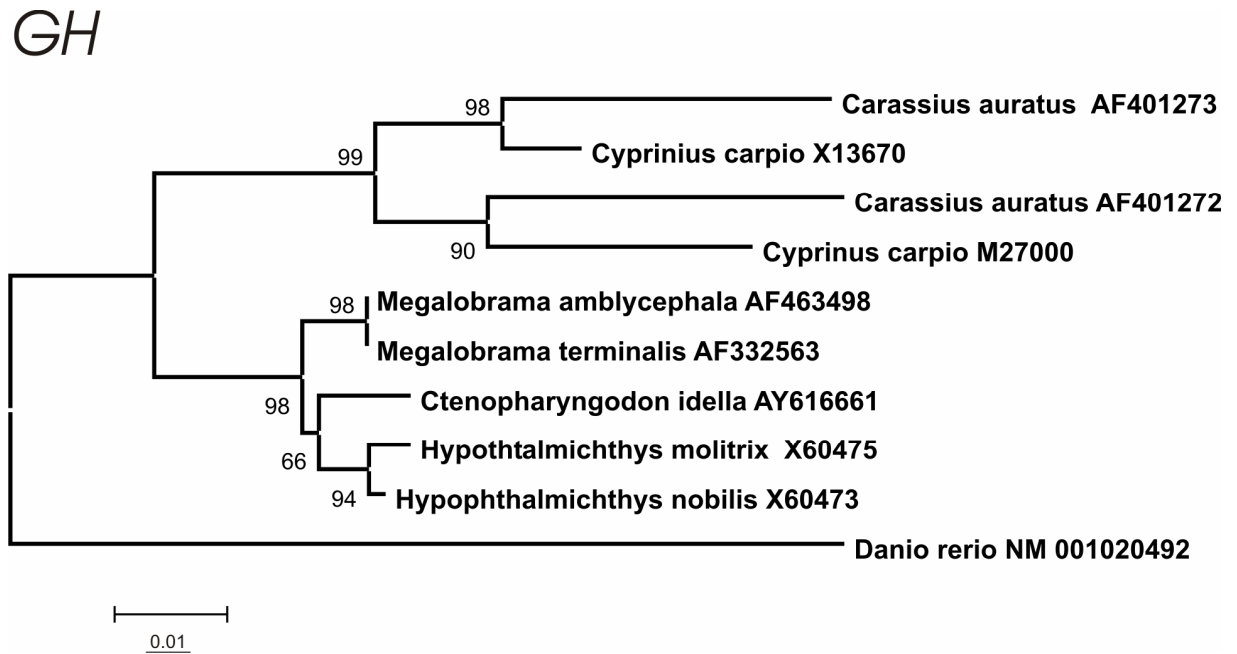


Figure III.1 Neighbor-Joining phylogenetic tree for growth hormone genes. Kimura's two-parameter distances are used.

The phylogenetic analysis has shown that pro-oncogene c - myc, gonadotropin releasing hormone (GRH), gonadotropin α , insulin like growth factor I (IGF - I) and insulin like growth factor I receptor genes have paralogues in common carp and in goldfish. The genes for pro-opiomelanocortin (POMC), growth hormone receptor (GHR), prolactin receptor and melanocortin 5 receptor have paralogues in common carp and one gene in goldfish. For the G-protein-coupled receptor (GPR34), Pit-1 (specific for pituitary transcription factor), integrin and ADP receptor-like GPCR, phylogenetic analysis has shown two genes for common carp but not any one for goldfish. The explanation for the lack of goldfish paralogues could be that not all genes or cDNA have been sequenced yet and thus are not available in GenBank.

The phylogenetic analysis of the above listed genes was done as described for growth hormone genes. The same tree topology was obtained with Neighbor-Joining method as well as with Maximum Likelihood methods for all of these genes. Therefore, only Neighbor-Joining trees are displayed.

The phylogenetic analysis of two proto-oncogene c - myc of common carp shows a similar tree topology as with GH genes (Fig. III.2). Each of two c - myc genes of common carp has one orthologous sequence from goldfish with 87% and 99% of bootstrap support.

Only one gene for zebrafish was found. For other cyprinids unfortunately no sequences were available in GenBank due to the incomplete nature of the database.

C-myc

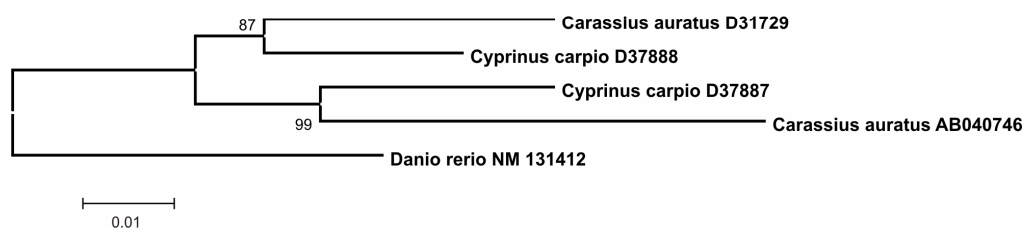


Figure III.2 Neighbor-Joining phylogenetic tree for c - myc. Kimura's two-parameter distances are used.

The phylogenetic trees for gonadotropin releasing hormone (GRH) and gonadotropin α also demonstrated two common carp and two goldfish genes, and one zebrafish or one zebrafish and one grass carp gene (Fig. III.3; Fig. III.4). In both gene trees one of paralogues of common carp grouped together with one of paralogues of goldfish, it means they are orthologues with each other. However, the second paralogue of common carp and goldfish lie in one group with paralogues from these species as well as with orthologous genes from other cyprinids. The phylogenetic trees of GRH and gonadotropin α have bootstrap support of 65% and 66%, respectively. The bootstrap support was higher (82-95%) if for gonadotropin α tree construction only goldfish paralogues or only common carp paralogues (Fig. III.4 B and C) as well as only first or second orthologues from both species (Fig. III.4 D and E) were used.

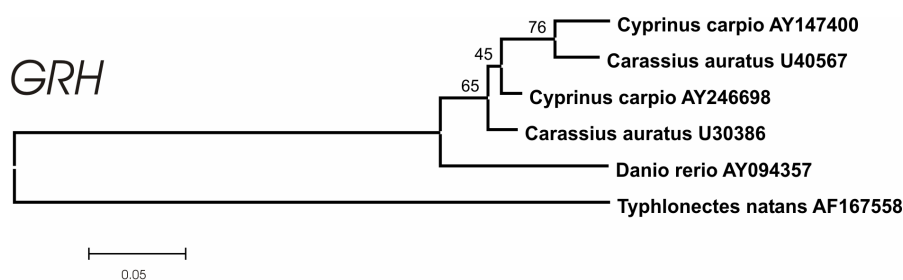


Figure III.3 Neighbor-Joining phylogenetic tree for Gonadotropin releasing hormone. Kimura's two-parameter distances are used.

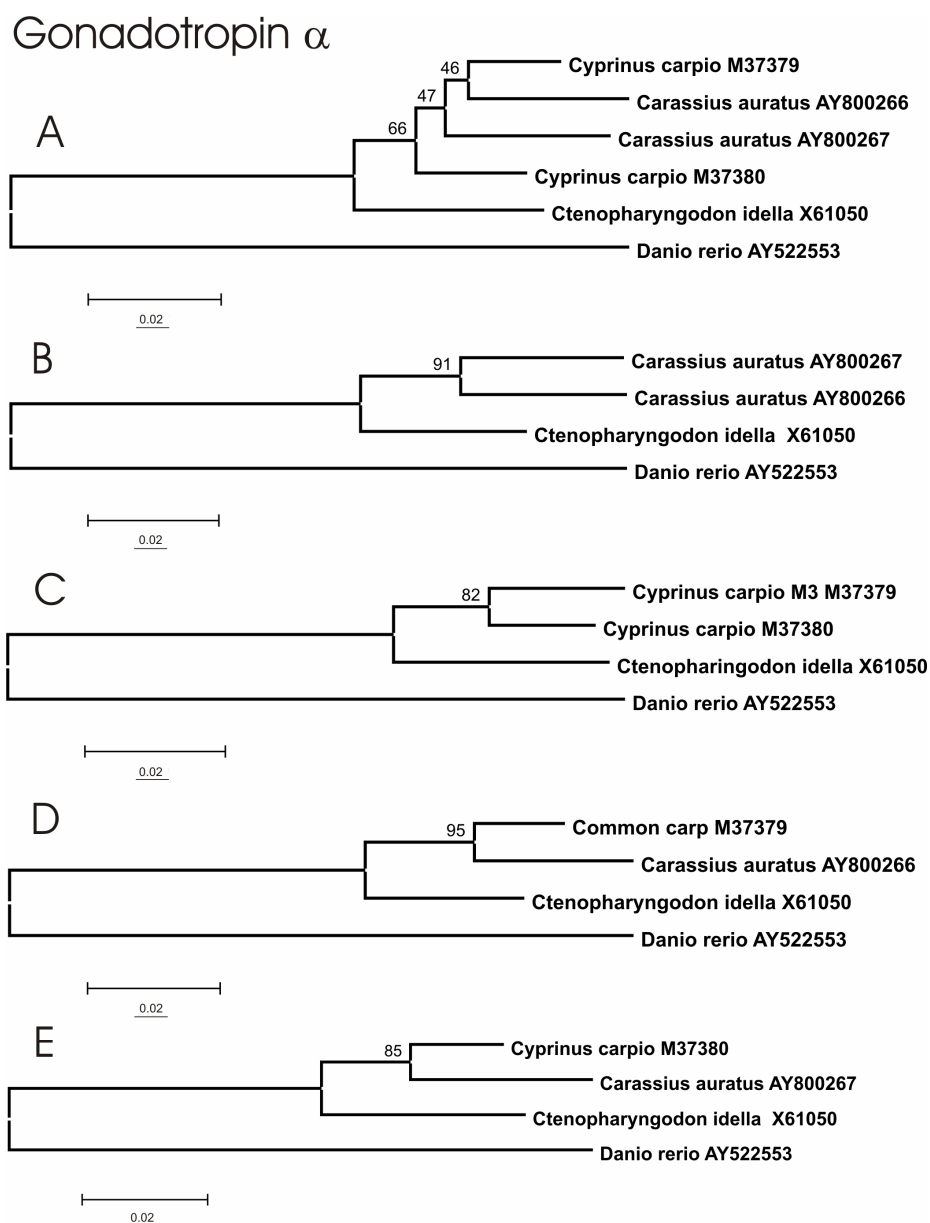


Figure III.4 Neighbor-Joining phylogenetic trees for Gonadotropin α . Kimura's two-parameter distances are used. A - two paralogues from common carp and two from goldfish; B - only paralogues from goldfish; C - only paralogues from common carp; D - first orthologues from common carp and goldfish; E - second orthologues from common carp and goldfish.

For POMC, GH receptor and prolactin receptor two genes for common carp and one for goldfish were found in GeneBank. One of the paralogues of common carp genes is a orthologous sequence to the single goldfish gene in phylogenetic analysis (Fig. III.5) with bootstrap support of 91% for POMC, 97% for GH receptor and 96% for prolactin receptor. It is possible that the second paralogue of these genes for goldfish was not sequenced yet and is not available in GenBank.

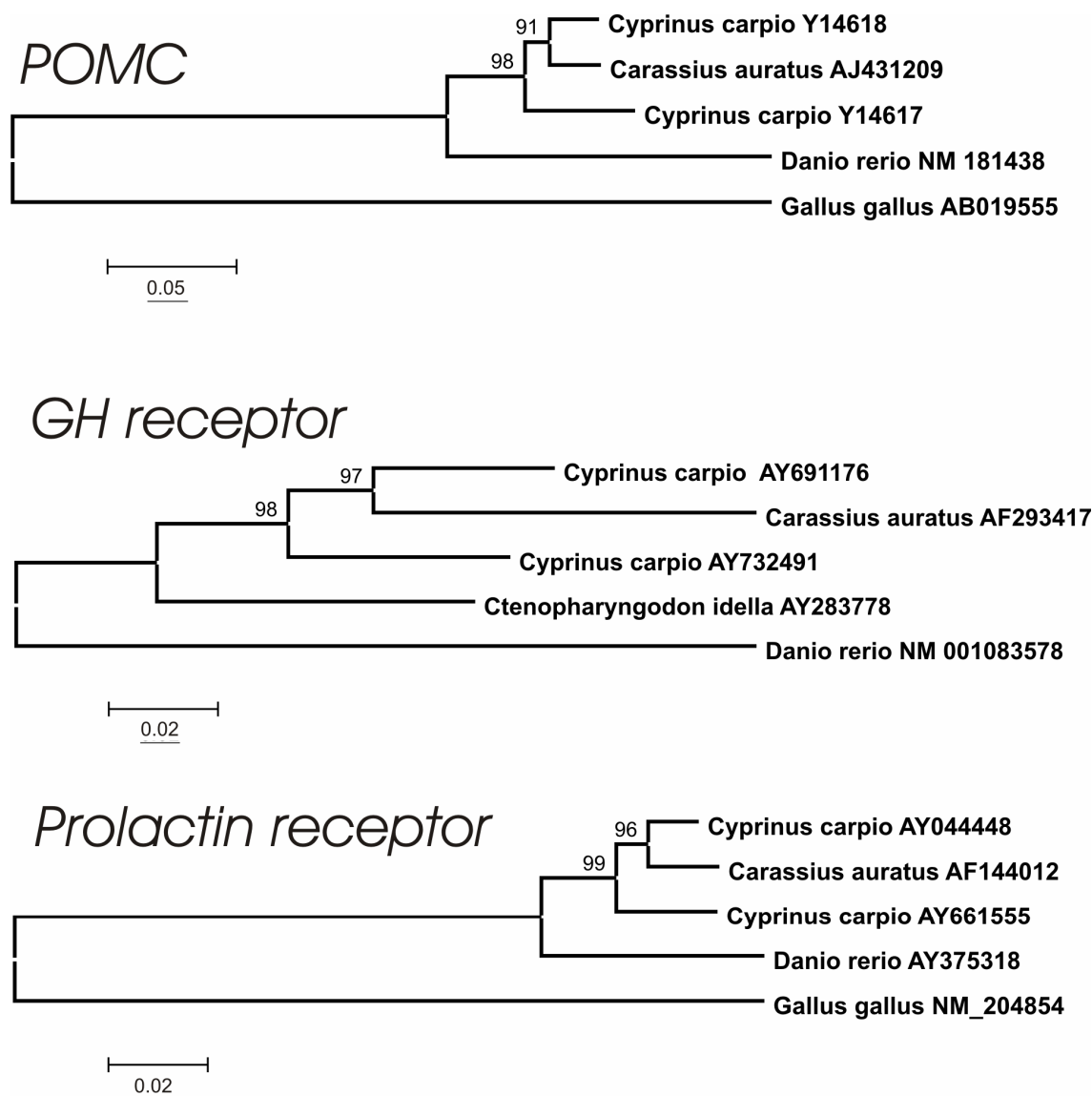


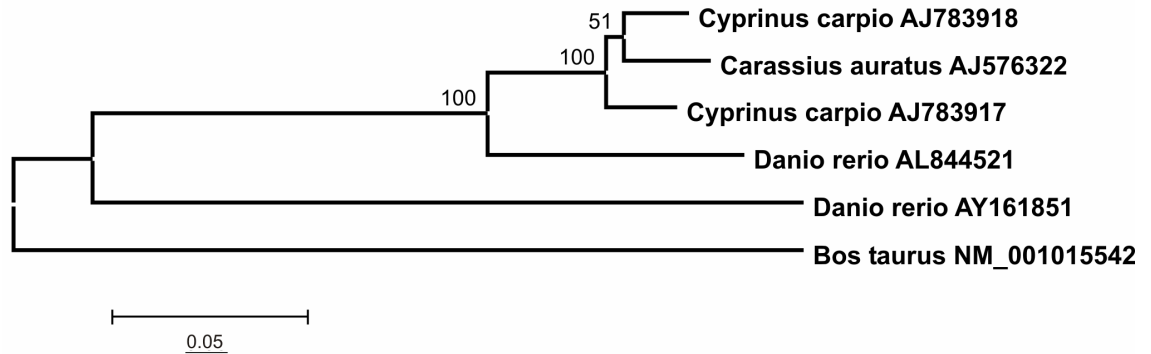
Figure III.5 Neighbor-Joining phylogenetic trees for POMC, GH receptor and Prolactin receptor. Kimura's two-parameter distances are used.

The tree topology of melanocortin 5 receptor revealed that there are two genes for zebrafish. Only one of them is in the same monophyletic group with two genes of common carp and one gene of goldfish (100% bootstrap support) (Fig. III.6). One of two paralogues of common carp is a orthologous sequence to the single sequence from goldfish. It could not be excluded that the second gene in goldfish was not sequenced up to now.

The phylogenetic analysis of G-protein-coupled receptor GPR34 showed two monophyletic groups (100% bootstrap support for each). The first monophyletic group has two paralogous common carp sequences and one orthologue to them from zebrafish. The

second monophyletic group has one common carp sequence and one from zebrafish (Fig. III.6).

Melanocortin 5 receptor



G-protein-coupled receptor GPR34

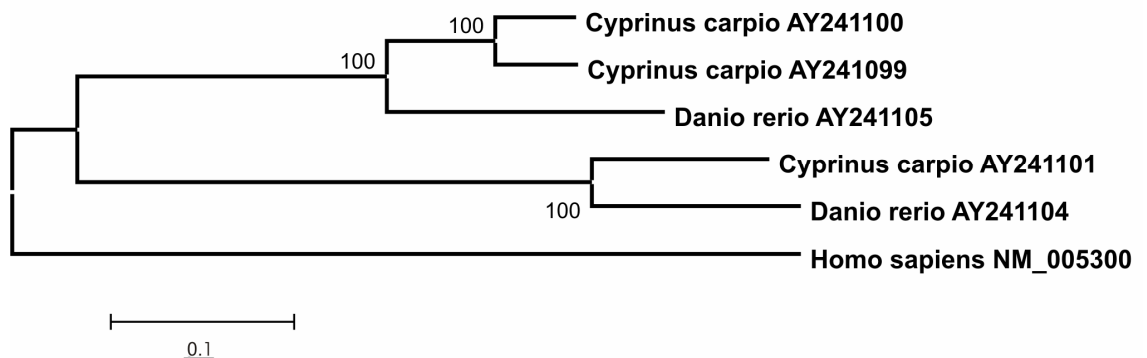
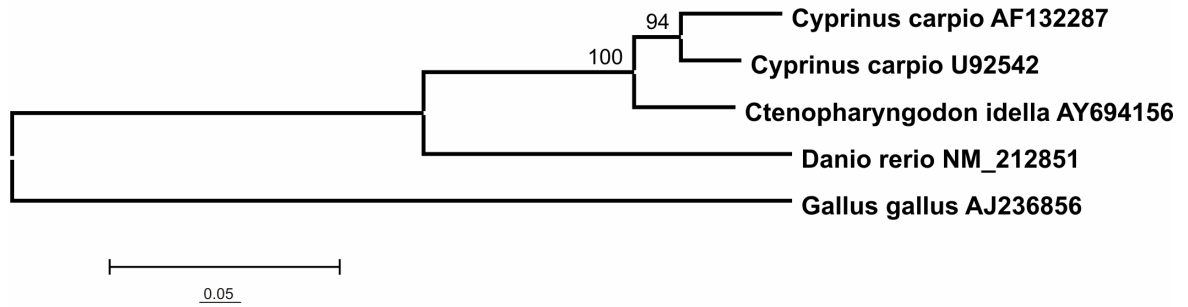


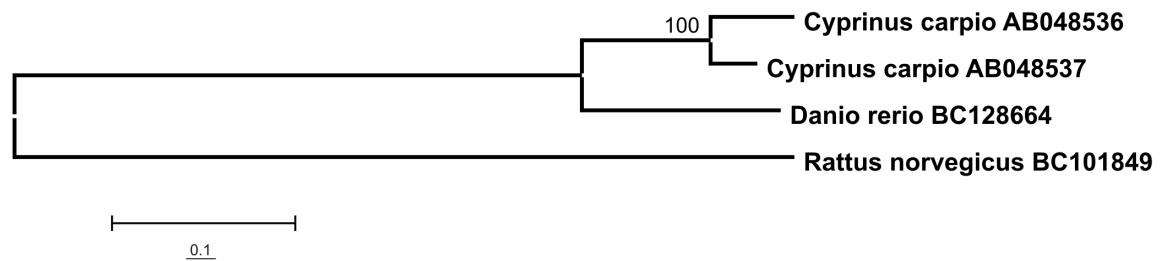
Figure III.6 Neighbor-Joining phylogenetic trees for Melanocortin 5 receptor, G-protein-coupled receptor GPR34. Kimura's two-parameter distances are used.

The phylogenetic analysis of genes for Pit-1, integrin and ADP receptor-like GPCR revealed that there are two genes in common carp but only one in zebrafish. The common carp genes are sister sequences. The bootstrap support is 94% for Pit-1, 100% for Integrin and 96% for ADP receptor paralogues (Fig. III.7). No goldfish sequences for these genes were available in GeneBank.

Pit-1



Integrin



ADP receptor

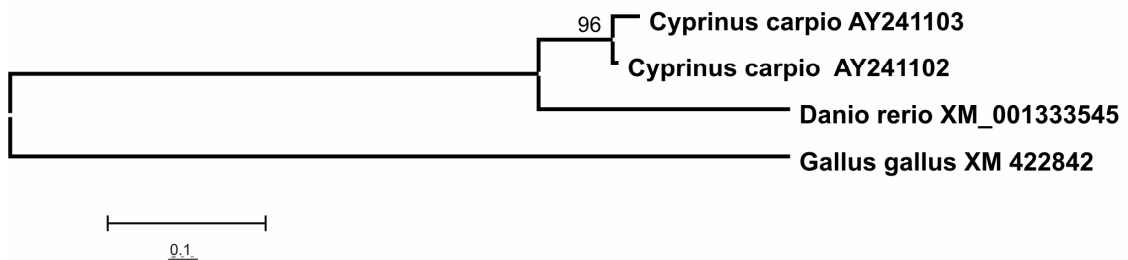


Figure III.7 Neighbor-Joining phylogenetic trees for Pit-1, Integrin and ADP receptor. Kimura's two-parameter distances are used.

Both common carp and goldfish have two genes for IGF - I as shown by the tree topology. These genes belong to a group with 53% bootstrap support. The paralogues of goldfish are grouped with each other and paralogous sequences from common carp form some “out group” to goldfish genes (Fig. III.8). In the other clade with 99% bootstrap support lie one gene from grass carp, from wuchang bream and from black Amur bream.

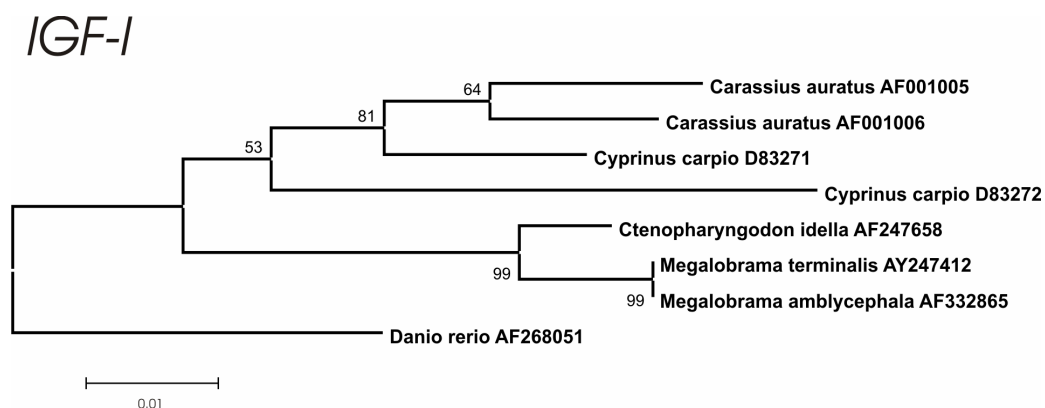


Figure III.8 Neighbor-Joining phylogenetic tree for IGF - I. Kimura's two-parameter distances are used.

The phylogenetic analysis of IGF - I receptor gene sequences of common carp, goldfish, diploid cyprinid species and fishes from other families gave the following results. There are two different monophyletic groups with common carp, goldfish and zebrafish gene sequences in each of them with 100% bootstrap support for each group (Fig. III.9). In addition, a third monophyletic group with 100% bootstrap support includes one gene from *Psetta maximus* and two from *Epinephelus coioides*.

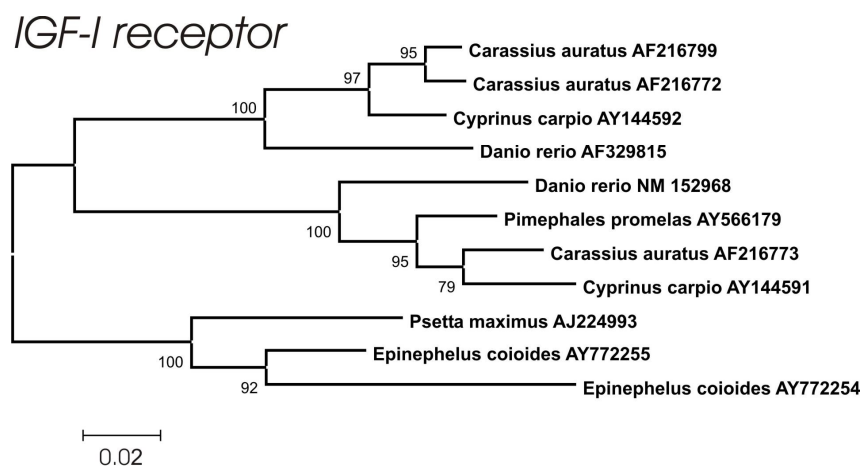


Figure III.9 Neighbor-Joining phylogenetic tree for IGF - I receptor. Kimura's two-parameter distances are used.

Rates of evolution and functional divergence of duplicated genes of common carp and goldfish were analysed. Groups of orthologous and paralogous genes analysed together are listed in Table III.2.

Table III-2 GenBank accession numbers for paralogous genes of common carp and goldfish that were compared with zebrafish orthologues in Tajima's test (1993). The same paralogues were analysed if non-synonymous mutations occur in such a way as to change protein hydrophobicity, charge or polarity to a greater extent than is expected under random substitution.

Gene	Species		
	Common carp I and II paralogues	Carassius auratus I and II paralogues	Danio rerio
Growth hormone	M27000 (cDNA) and X13670 (cDNA)	AF401272 (cDNA) and AF401273 (cDNA)	NM 001020492 (cDNA)
Insuline like growth factor I (IGF - I)	D83272 (cDNA) and D83271 (cDNA)	-	AF268051 (cDNA)
Growth hormone receptor (GHR)	AY732491 (cDNA) and AY691176 (cDNA)	-	NM 001083578 (cDNA)
Pit 1	AF132287 (gene) and U92542 (gene)	-	NM_212851 (cDNA)
Gonadotropin releasing hormone (GRH)	AY147400 (cDNA) and AY246698 (gene)	U40567 (cDNA) and U30386 (cDNA)	AY094357 (cDNA)
Gonadotropin alpha subunit	M37379 (cDNA) and M37380 (cDNA)	AY800266 (cDNA) and AY800267 (cDNA)	AY522553 (cDNA)
Prolactin receptor	AY044448 (cDNA) and AY661555 (cDNA)	-	AY375318 (cDNA)
ADP receptor - like GPCR	AY241103 (gene) and AY241102 (gene)	-	XM_001333545(cDNA)
c - myc	D37888 (gene) and D37887 (gene)	D31729 (gene) and AB040746 (gene)	NM 131412 (cDNA)
Melanocortin 5 receptor	AJ783918 (cDNA) and AJ783917 (cDNA)	-	AL 844521 (gene)
G-protein-coupled receptor GPR34	AY241100 (cDNA) and AY241099 (cDNA)	-	AY241105 (cDNA)
POMC	Y14618 (cDNA) and Y14617 (cDNA)	-	NM 181438 (cDNA)
Integrine	AB048536 (cDNA) and AB048537 cDNA)	-	BC 128664 (cDNA)

Tables III.3 and III.4 show mean dS/100 and dN/100 sites in comparison between duplicated common carp genes and in comparison between duplicated goldfish genes. In addition, mean dS/100 and dN/100 sites in comparison between each of two paralogues from common carp or goldfish and one orthologous zebrafish sequence are presented in these tables.

The synonymous nucleotide substitution rate among 13 duplicated gene pairs of common carp (Tab. III.3) ranged from 8 ± 1.9 to 25.4 ± 3.7 substitutions/100 sites and the average was 16 ± 3.25 substitutions/100 sites. Among 5 duplicated goldfish gene pairs (Tab. III.4) the synonymous nucleotide substitution rate varied from 9.9 ± 3.1 to 34.5 ± 4.7 substitutions/100 sites and the average was 21.4 ± 5.1 substitutions/100 sites. The range of synonymous nucleotide substitution rates for the studied paralogues from common carp and goldfish show that gene duplications could occur roughly at the same time.

The non-synonymous nucleotide substitution rate among 13 duplicated gene pairs of common carp (Tab. III.3) ranged from 0.1 ± 0.1 to 8.7 ± 1.1 substitutions/100 sites and the average was 3.8 ± 0.76 substitutions/100 sites. Among 5 duplicated goldfish gene pairs (Tab. III.4) the non-synonymous nucleotide substitution rate varied from 1.5 ± 0.7 to 6.3 ± 1.7 substitutions/100 sites and the average was 3.54 ± 0.96 substitutions/100 sites.

For all genes, dS exceeded dN in all comparisons, as would be expected for genes under purifying selection (Nei, 1987).

The divergence time was calculated for common carp duplicated genes and goldfish duplicated genes using the average synonymous nucleotide substitution rates. Divergence time for 13 duplicated genes of common carp was 17 Myr and for 5 duplicated genes of goldfish 23.2 Myr.

The relative test of evolution rate (Tajima, 1993; Nei & Kumar, 2000) showed statistically significant increase in rate of evolution in one of two growth hormone, gonadotropin releasing hormone and *IGF - I* genes of common carp (Tab. III.5 and Tab. III.7) and in one of two gonadotropin releasing hormone and c - myc genes of goldfish (Tab. III.6 and Tab. III.8).

One of the common carp GH paralogues (*GH I*) has increased evolution rate both on the amino acid level ($p=0.008$) and on the basis of all codon positions ($p=0.041$) (at most at first codon position: $p=0.008$). One of two common carp *GRH* genes (I) shows increased rate on the basis of all codon positions ($p=0.005$) (at most at third: $p=0.020$) and on the amino acid level it is nearly significant ($p=0.059$). One of the duplicated *IGF - I* genes of common carp (I) has nearly significantly increased evolution rate on nucleotide level (at most at second codon position: $p=0.046$) and significantly increased on the amino acid level

($p=0.008$) (Tab. III.5 and Tab. III.7). An increased evolution rate both on the amino acid level ($p=0.003$) and on the basis of all codon positions ($p=0.0006$) is found in one of two goldfish *c - myc* genes (II). The *GRH* genes of goldfish also show different evolution rate. One of them (I) has increased rate on the basis of all codon positions ($p=0.001$) (at most at third: $p=0.005$) and on the amino acid level it is nearly significant ($p=0.058$) (Tab. III.6 and Tab. III.8).

The evolution rates between the orthologous genes of common carp and goldfish are presented in Tab. III.9. The *GH II* orthologous genes show differences in evolution rate on the nucleotide level. The number of unique nucleotide substitutions is significantly higher in the *GH II* gene of goldfish than of common carp ($p=0.012$). The *c - myc II* genes have different evolution rates both at nucleotide ($p=0.0001$) and amino acid level ($p=0.006$). The *c - myc II* gene of goldfish evolves faster.

The genes were analysed for positive Darwinian selection ($pNR > pNC$) regarding charge, polarity and hydrophobicity (Tab. III.10). None of them showed evidence for relaxation of functional constraints or positive Darwinian selection for hydrophobicity. This provides strong evidence for purifying selection that is keeping properties such as hydrophobicity.

The *GH* genes of common carp show pNR/pNC values close to 1 for polarity, which would point to neutrality. Neutrality would imply relaxation of functional constraints in one of these genes or at least of part of the gene. On the other hand, the low ratios of $pNR/pNC < 1$ for goldfish *GH* paralogues seem to suggest that structural constraints keep them from changing their polarity.

The *c - myc*, *GH* receptor, *Integrin*, *POMC* genes of common carp and *c - myc* of goldfish (Tab. III.10) also show relaxation of functional constraints for polarity in one of their paralogues. The *GRH* and *Pit-1* genes in common carp as well as *GRH* in goldfish have pNR/pNC ratio > 1 for change of polarity what is evidence for positive Darwinian selection in one of the paralogues. In addition, the *Pit-1* genes of common carp show positive Darwinian selection in one of the paralogues for charge. Positive selection for charge was also found in one of the paralogues for gonadotropin α in goldfish. One of the *POMC* genes of common carp shows also relaxation of functional constraints for charge.

Table III-3 Number of synonymous nucleotide substitutions per 100 synonymous sites ($d_S/100$) and number of non-synonymous nucleotide substitutions per 100 non-synonymous sites $d_N/100$ (+ SE) between two genes of common carp and between each of two duplicated sequences of common carp and an orthologous zebrafish sequence.

Gene	Common carp I vs Common carp II			Mean $d_N \pm SE$		Mean $d_S \pm SE$	
	Mean $d_S \pm SE$	Mean $d_N \pm SE$	d_N/d_S	Common carp I and zebrafish	Common carp II and zebrafish	Common carp I and zebrafish	Common carp II and zebrafish
Growth hormone (GH)	16.3 \pm 3.7	1.5 \pm 0.5	0.09	5.4 \pm 1.1	3.9 \pm 0.9	61.0 \pm 10.3	56.1 \pm 9.5
Growth hormone receptor (GHR)	25.1 \pm 4.1	4.2 \pm 0.9	0.16	14.1 \pm 1.8	14.8 \pm 2.0	52.2 \pm 7.2	57.9 \pm 8.6
PIT1	13.1 \pm 3.1	1.3 \pm 0.5	0.1	3.7 \pm 0.9	3.2 \pm 0.8	77.4 \pm 13.9	72.7 \pm 12.7
Gonadotropin releasing hormone (GRH)	15.3 \pm 5.7	3.6 \pm 1.4	0.24	8.8 \pm 2.2	6.0 \pm 2.0	61.3 \pm 16.9	40.0 \pm 11.1
Gonadotropin alpha subunit	8.0 \pm 3.5	2.6 \pm 0.9	0.33	13.5 \pm 2.7	12.4 \pm 2.6	38.7 \pm 9.2	34.7 \pm 8.2
Prolactin receptor	22.8 \pm 2.7	7.0 \pm 0.7	0.31	10.6 \pm 1.0	10.1 \pm 1.0	71.8 \pm 6.8	64.4 \pm 5.9
ADP receptor - like GPCR	8.0 \pm 1.9	4.8 \pm 1.0	0.6	11.4 \pm 1.5	11.0 \pm 1.5	65.3 \pm 8.7	64.2 \pm 8.5

Table III-3 continued

Gene	Common carp I vs Common carp II			Mean d _N ± SE		Mean d _S ± SE	
	Mean d _S ± SE	Mean d _N ± SE	d _N /d _S	Common carp I and zebrafish	Common carp II and zebrafish	Common carp I and zebrafish	Common carp II and zebrafish
c - myc	25.4 ± 3.7	2.4 ± 0.5	0.09	3.6 ± 0.6	3.2 ± 0.6	32.9 ± 4.2	38.7 ± 5.1
Melanocortin 5 receptor	15.2 ± 2.7	0.1 ± 0.1	0.007	1.9 ± 0.5	1.6 ± 0.5	56.8 ± 7.5	50.4 ± 6.5
G-protein-coupled receptor GPR34	12.4 ± 2.3	8.7 ± 1.1	0.7	19.7 ± 1.9	21.2 ± 1.9	50.7 ± 6.8	53.1 ± 6.9
POMC	17.3 ± 3.8	4.4 ± 1.0	0.25	10.1 ± 1.6	10.6 ± 1.6	54.5 ± 8.5	59.5 ± 9.6
Integrine	12.9 ± 1.4	6.9 ± 0.6	0.53	18.4 ± 1.0	19.1 ± 1.0	44.4 ± 3.2	45.2 ± 3.2
Insuline like growth factor (IGF - I)	15.9 ± 3.7	1.9 ± 0.7	0.06	3.4 ± 0.9	1.4 ± 0.6	23.6 ± 5.0	21.3 ± 4.6
Average	16 ± 3.25	3.8 ± 0.76		9.6 ± 1.4	9.1 ± 1.3	53.1 ± 8.3	50.6 ± 7.7

*The two IGF - I receptor genes in common carp were the result of an ancient duplication (see phylogenetic tree, Fig.III.9) prior to the tetraploidization event in this fish species. Therefore, these genes were not included in Tab. III.3.

Table III-4 Number of synonymous nucleotide substitutions per 100 synonymous sites ($d_S/100$) and number of non-synonymous nucleotide substitutions per 100 non-synonymous sites $d_N/100$ (+ SE) between two genes of goldfish and between each of two duplicated sequences of goldfish and an orthologous zebrafish sequence.

Gene	Goldfish I vs Goldfish II			Mean d _N ± SE		Mean d _S ± SE	
	Mean d _S ± SE	Mean d _N ± SE	d _N /d _S	Goldfish I and zebrafish	Goldfish II and zebrafish	Goldfish I and zebrafish	Goldfish II and zebrafish
Growth hormone (GH)	27.3 ± 5.6	3.8 ± 0.9	0.14	4.9 ± 1.1	4.3 ± 1.1	58.7 ± 10.3	72.1 ± 13.2
Gonadotropin releasing hormone (GRH)	17.1 ± 6.2	6.3 ± 1.7	0.37	11.1 ± 2.7	7.2 ± 2.1	53.3 ± 15.2	31.6 ± 9.3
Gonadotropin alpha subunit	18.2 ± 5.9	1.5 ± 0.7	0.08	12.3 ± 2.5	11.5 ± 2.5	44.3 ± 10.1	53.4 ± 13.0
c - myc	34.5 ± 4.7	4.4 ± 0.7	0.13	3.0 ± 0.6	5.2 ± 0.8	39.6 ± 4.9	45.5 ± 5.6
Insuline like growth factor I (IGF - I)	9.9 ± 3.1	1.7 ± 0.8	0.17	3.7 ± 1.1	3.1 ± 1.0	19.2 ± 4.5	21.2 ± 4.7
Average	21.4 ± 5.1	3.54 ± 0.96		7.0 ± 1.6	6.3 ± 1.5	43.0 ± 9.0	44.8 ± 9.16

*The two IGF - I receptor genes in goldfish were the result of an ancient duplication (see phylogenetic tree Fig. III.9) prior to the tetraploidization event in this fish species. Therefore, these genes were not included in Tab. III.4.

Table III-5 Results of the nonparametric relative rate test of Tajima (1993) comparing each of two common carp paralogues with their zebrafish orthologue. Substitutions were computed together for first, second and third codon position and for amino acids.

Gene	Common carp duplicated genes						
	Nucleotide sequence			Amino acid sequence			
	Sites	m1 ^a	m2 ^b	χ^2	significant ^c	Sites	m1 ^a m2 ^b χ^2 significant ^c
Growth hormone (GH)	633	17	7	4.17	p=0.041 Yes	210	7 0 7 p=0.008 Yes
Insuline like growth factor (IGF - I)	486	16	7	3.52	p=0.061 No	161	7 0 7 p=0.008 Yes
Growth hormone receptor (GHR)	736	24	32	1.14	p=0.285 No	245	9 8 0.06 p=0.808 No
Pit 1	667	13	8	1.19	p=0.275 No	221	3 1 1 p=0.317 No
Gonadotropin releasing hormone(GRH)	261	13	2	8.07	p=0.005 Yes	86	6 1 3.57 p=0.059 No
Gonadotropin alpha subunit	354	8	4	1.33	p=0.248 No	117	3 2 0.2 p=0.655 No
Prolactin receptor	1779	82	63	2.49	p=0.115 No	593	30 30 0 p=1.000 No
ADP receptor - like GPCR	914	9	6	0.6	p=0.439 No	304	4 1 1.8 p=0.180 No
c – myc	1185	32	38	0.51	p=0.473 No	394	10 8 0.22 p=0.637 No
Melanocortin 5 receptor	987	18	9	3	p=0.083 No	328	1 0 1 p=0.317 No
G-protein-coupled receptor GPR34	1069	36	47	1.46	p=0.227 No	355	17 21 0.42 p=0.516 No
POMC	669	14	20	1.06	p=0.303 No	222	6 7 0.08 p=0.782 No
Integrine	3420	103	121	1.45	p=0.229 No	1139	41 53 1.53 p=0.216 No

See also footnote for Table III.6.

Table III-6 Results of the nonparametric relative rate test of Tajima (1993) comparing each of two goldfish paralogues with their zebrafish orthologue. Substitutions were computed together for first, second and third codon position and for amino acids.

Gene	Goldfish duplicated genes									
	Nucleotide sequence					Amino acid sequence				
	Sites	m1 ^a	m2 ^b	χ^2	Significant ^c	Sites	m1 ^a	m2 ^b	χ^2	Significant ^c
Growth hormone (GH)	567	18	23	0.61	p=0.435 No	188	6	9	0.6	p=0.439 No
Insuline like growth factor (IGF - I)	486	8	8	0	p=1.00 No	161	3	2	0.2	p=0.655 No
Gonadotropin releasing hormone (GRH)	261	16	2	10.07	p=0.001 Yes	86	8	2	3.6	p=0.058 No
Gonadotropin alpha subunit	351	7	9	0.25	p=0.617 No	116	3	1	1	p=0.317 No
c - myc	1167	38	66	7.54	p=0.0006 Yes	388	7	23	8.53	p=0.003 Yes

Genes that show a statistically significant increase in rate of evolution in one of duplicates are indicated in bold. Genes that show a nearly statistically significant increase in rate of evolution in one of duplicates are indicated in blue. ^am1 is the number of unique substitutions in paralogue 1; ^bm2 is the number of unique substitutions in paralogue 2; ^csignificant at the 95% confidence level (p<0.05).

Table III-7 Results of the nonparametric relative rate test of Tajima (1993) comparing each of two common carp paralogues with their zebrafish orthologue. Substitutions were computed separately for first, second and third codon positions.

Codon position Gene	Common carp duplicated genes														
	1st					2nd					3rd				
	Sites	m1 ^a	m2 ^b	χ^2	Significant ^c	Sites	m1 ^a	m2 ^b	χ^2	Significant ^c	Sites	m1 ^a	m2 ^b	χ^2	Significant ^c
Growth hormone (GH)	211	7	0	7	p=0.008	211	3	0	3	p=0.083	211	7	7	0	p=1.00
Insuline like growth factor (IGF - I)	162	8	6	0.29	p=0.593	161	4	0	4	p=0.046	163	4	1	1.8	p=0.180
Growth hormone receptor (GHR)	564	5	6	0.09	p=0.763	564	5	6	0.09	p=0.763	564	17	21	0.42	p=0.516
Pit 1	222	1	1	0	p=1.0	223	1	0	1	p=0.317	222	11	7	0.89	p=0.346
Gonadotropin releasing hormone (GRH)	87	3	1	1	p=0.317	87	2	0	2	p=157	87	8	1	5.44	p=0.020
Gonadotropin alpha subunit	118	2	2	0	p=1.00	118	1	0	1	p=0.317	118	5	2	1.29	p=0.257
Prolactin receptor	593	20	14	1.06	p=0.303	593	16	18	0.12	p=0.732	593	46	31	2.92	p=0.087
ADP receptor - like GPCR	306	3	1	1	p=0.317	305	3	2	0.2	p=0.655	306	3	2	0.2	p=655
c – myc	395	5	5	0	p=1.00	395	5	5	0	p=1.00	395	22	29	0.96	p=0.327
Melanocortin 5 receptor	329	2	1	0.33	p=0.564	329	0	0			329	16	8	2.67	p=0.102
G-protein-coupled receptor GPR34	356	15	17	0.13	p=0.724	356	8	14	1.64	p=0.201	357	13	16	0.31	p=0.577
POMC	223	6	11	1.47	p=0.225	224	5	7	0.33	p=0.564	222	3	2	0.2	p=0.655
Integrine	1140	26	40	2.97	p=0.085	1140	26	26	0	p=1.00	1140	51	55	0.15	p=0.698

Genes that show statistically significant increase in rate of evolution in one of duplicates are indicated in bold. ^am1 is the number of unique substitutions in paralogue 1; ^bm2 is the number of unique substitutions in paralogue 2; ^csignificant at the 95% confidence level (p<0.05).

Table III-8 Results of the nonparametric relative rate test of Tajima (1993) comparing each of two goldfish paralogues with their zebrafish orthologue. Substitution were computed separately for first, second and third codon positions.

Codon positon		Goldfish duplicated genes													
		1st				2nd				3rd					
Gene	Sites	m1 ^a	m2 ^b	χ^2	Significant ^c	Sites	m1 ^a	m2 ^b	χ^2	Significant ^c	Sites	m1 ^a	m2 ^b	χ^2	Significant ^c
Growth hormone (GH)	189	6	2	2	p=0.157	189	3	4	0.14	p=0.705	189	9	17	2.46	p=0.117
Insuline like growth factor (IGF - I)	162	5	6	0.09	p=0.763	161	2	1	0.33	p=0.564	163	1	1	0	p=1.00
Gonadotropin releasing hormone (GRH)	87	2	1	0.33	p=0.564	87	6	1	3.57	p=0.059	87	8	0	8	p=0.005
Gonadotropin alpha subunit	117	2	2	0	p=1.00	117	0	0			117	5	7	0.33	p=0.564
c - myc	389	4	11	3.27	p=0.071	389	6	14	3.2	p=0.074	389	28	41	2.45	p=0.118

Genes that show statistically significant increase in rate of evolution in one of duplicates are indicated in bold. Genes that show a nearly statistically significant increase in rate of evolution in one of duplicates are indicated in blue. ^am1 is the number of unique substitutions in paralogue 1; ^bm2 is the number of unique substitutions in paralogue 2; ^csignificant at the 95% confidence level (p<0.05).

Table III-9 Results of the nonparametric relative rate test of Tajima (1993) comparing orthologues I or II from common carp and goldfish (orthologues were determined in phylogenetic analysis).

Gene	Common carp and goldfish duplicated genes I				
	Nucleotide sequence		Amino acid sequence		
	Sites	m1 ^a	m2 ^b	χ^2	Significant ^c
Growth hormone (GH)	567	13	12	0.04	p=0.841 No
Gonadotropin releasing hormone (GRH)	261	4	6	0.4	p=0.527 No
Gonadotropin alpha subunit	351	3	7	1.6	p=0.206 No
c - myc	1191	25	31	0.64	p=0.423 No
Common carp and goldfish duplicated genes II					
Gene	Nucleotide sequence		Amino acid sequence		
	Sites	m1 ^a	m2 ^b	χ^2	Significant ^c
Growth hormone (GH)	567	4	15	6.37	p=0.012 Yes
Gonadotropin releasing hormone (GRH)	261	4	3	0.14	p=0.705 No
Gonadotropin alpha subunit	354	4	10	2.57	p=0.109 No
c - myc	1173	23	51	10.59	p=0.0001 Yes

Genes that show a statistically significant increase in rate of evolution in one of duplicates are indicated in bold. Substitutions were computed for first, second and third codon position together and for amino acids. ^am1 is the number of unique substitutions in the common carp orthologue; ^bm2 is the number of unique substitutions in the goldfish orthologue; ^csignificant at the 95% confidence level (p<0.05).

Table III-10 Ratio of radical and conservative amino acid changes in common carp and goldfish paralogues.

Gene	Common carp				Goldfish				Common carp				Goldfish			
	Polarity		Polarity		Polarity		Polarity		Charge		Charge		Charge		Charge	
	p _{NC}	p _{NR}	p _{NR} /p _{NC}	p _{NC}	p _{NR}	p _{NR} /p _{NC}	p _{NC}	p _{NR}	p _{NR} /p _{NC}	p _{NC}	p _{NR}	p _{NR} /p _{NC}	p _{NC}	p _{NR}	p _{NR} /p _{NC}	p _{NR} /p _{NC}
GH	0.0135	0.0141	1.04	x	x	x	x	x	x	x	x	x	x	x	x	x
c - myc	0.0239	0.0233	0.98	0.0419	0.0432	1.03	x	x	x	x	x	x	x	x	x	x
GRH	0.0152	0.0726	4.78	0.0461	0.087	1.89	x	x	x	x	x	x	x	x	x	x
GH receptor	0.835	0.913	1.093	0	0	0	x	x	x	x	x	x	0	0	0	0
Integrin	0.463	0.515	1.112	0	0	0	x	x	x	x	x	x	0	0	0	0
POMC	0.414	0.447	1.06	0	0	0	0.414	0.439	1.07	0	0	0	0	0	0	0
Pit 1	0.0381	0.048	1.26	0	0	0	0.0358	0.0526	1.47	0	0	0	0	0	0	0
IGF - I	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
Prolactin receptor	x	x	x	0	0	0	x	x	x	x	x	x	0	0	0	0
Melanocortin 5 receptor	x	x	x	0	0	0	x	x	x	x	x	x	0	0	0	0
Gonadotropin α	x	x	x	x	x	x	x	x	x	x	x	x	0	0	0	0
Melanocortin 5 receptor	x	x	x	0	0	0	x	x	x	0.0121	0.0189	1.56	0	0	0	0

p_{NC} is the proportion of conservative non-synonymous differences per conservative non-synonymous site; p_{NR} is the proportion of radical (non-conservative) non-synonymous differences per radical non-synonymous site; Genes that show positive Darwinian selection (ratio p_{NR} : p_{NC} > 1) indicated in bold; Genes that show purifying selection (ratio p_{NR} : p_{NC} < 1) indicated with X; Genes for which only one paralogue was available in GeneBank indicated with 0

III.4 Discussion

The *GH*, *c - myc*, *POMC*, GH receptor, prolactin receptor and melanocortin 5 receptor phylogenetic tree topologies demonstrate that gene duplications occurred in common carp and goldfish before they separated from each other (Fig. III.1, Fig. III.2, Fig. III.5, Fig. III.6). In view of that common carp and goldfish have the same tetraploid ancestor, the duplication of these genes arose most probably through whole genome duplication. In the present study for *POMC*, growth hormone receptor and prolactin receptor two cDNA sequences were found for two genes from common carp but only one from goldfish. It could be explained by the incomplete nature of GenBank. On the other hand, one of the paralogues of these genes could be silenced in goldfish. The phylogenetic trees of gonadotropin α and *GRH* (Fig. III.3 and Fig. III.4) show that one paralogue from common carp and one from goldfish are orthologues with each other. The second paralogue from common carp and the second paralogue from goldfish form the outgroup to other paralogues from goldfish and common carp. They lie in one group with paralogues from these species as well as with orthologous genes from other cyprinids. The “outgroup topology” for gonadotropin α and *GRH* are poorly supported by bootstrap reiteration. It could be that outgroup topologies are reconstruction artefacts, perhaps caused by unequal rates of evolution in paralogues of common carp and/or goldfish. The Tajima’s relative rate test has shown that duplicated *GRH* genes in common carp as well as in goldfish have different evolution rates and comparison of p_{NR} and p_{NC} values has demonstrated positive Darwinian selection in one paralogue of common carp as well as of goldfish (Tab. III.10). In case of gonadotropin α differences in evolution between goldfish paralogues were also found. The p_{NR}/p_{NC} ratio was >1 for charge what also demonstrates positive Darwinian selection in one of the paralogues (Tab. III.10).

The dendrograms for G-protein-coupled receptor CPR34, *Pit-1*, Integrine, ADP receptor-like GPCR also show two sister sequences of common carp for one zebrafish sequence (Fig. III.6 and Fig. III.7). Because common carp is tetraploid the paralogues of these genes most probably arose through whole genome duplication.

Unfortunately, there are no completed gene linkage maps for common carp and goldfish. Therefore, it is not possible to proof if all studied duplicated genes are unlinked.

The tree topology of *IGF - I* showed that paralogues of goldfish are grouped with each other and paralogous sequences from common carp form some “out group” to goldfish genes (Fig. III.8). This could be explained by gene duplication that was lineage specific

separately for common carp and separately for goldfish. On the other hand, it could also be a tree reconstruction artefact caused by different evolution rates in *IGF - I* duplicated genes of common carp (Tab. III.5 and Tab. III.7).

The phylogenetic analysis of IGF - I receptor gene sequences showed two different monophyletic groups with one common carp, one goldfish and one zebrafish gene sequences in each of them. This tree topology demonstrates that this gene duplication occurred before divergence of zebrafish and ancestor of common carp and goldfish. Because there is only one gene from common carp and goldfish in each monophyletic group it could also be possible that some duplicated genes for IGF - I receptor were silenced in common carp and goldfish after the lineage specific genome duplication in ancestor of common carp and goldfish or that cDNA of them is not sequenced yet.

The dendrograms for Melanocortin 5 receptor and G-protein-coupled receptor GPR34 also show the ancient gene duplication prior to the tetraploidization event in common carp and goldfish ancestor.

Recently, it has been suggested that the large number of fish species (about 25,000; Nelson, 1994) and their tremendous morphological diversity might be due to a genome duplication event specific to the teleost lineage that occurred between 300 and 450 Myr ago (Amores et al., 1998; Wittbrodt et al., 1998; Meyer & Schartl, 1999; Taylor et al., 2001). Therefore, the additional gene duplication that is shown by the phylogenetic trees for IGF - I receptor, Melanocortin 5 receptor and G-protein-coupled receptor GPR 34 could possibly be arisen through genome duplication before radiation of teleosts. On the other hand, there is the third monophyletic group with 100% bootstrap support in the IGF - I receptor phylogenetic tree with one gene from fish *Psetta maximus* and two from *Epinephelus coioides*. In the future, more precisely studies are needed to understand if these gene duplications were caused by whole genome duplication in the teleost ancestor or if there were lineage specific gene duplications.

The average synonymous nucleotide substitution rate for 13 duplicated genes in common carp was 16 ± 3.25 substitutions/100 sites and was below than the average for 5 duplicated genes in goldfish (21.4 ± 5.1 substitutions/100 sites) (Tab. III.3 and Tab. III.4). This difference could be caused by the low number of genes studied from goldfish because more sequences are not available in GenBank. On the other hand, it could also be explained by some shorter generation interval in goldfish than in common carp (Wu & Li, 1985).

Divergence time calculated in the present study on the basis of duplicated genes from common carp was 17 Myr and on the basis of duplicated genes from goldfish 23.2 Myr. That is within the period of time that was calculated by other scientists (Risinger &

Larhammar, 1993; Larhammar & Risinger, 1994; David et al., 2003). The difference in divergence time between common carp and goldfish genes is coupled with the divergence of the average synonymous nucleotide substitution rates (see above).

Allotetraploidy of the common carp and goldfish ancestor might have confounded our ability to date the genome duplication. In autotetraploidy “duplicate” genes come from two individuals of the same species and are identical or are alleles at a given locus. With allotetraploidy the two genomes involved come from different species and may have diverged extensively at the faster evolving loci before the tetraploidy, i.e. duplication event (Spring, 1997). Thus, for genome duplication via allotetraploidy, difference between coorthologues begins before the tetraploidy event (i.e. genome duplication).

For all duplicated gene pairs in the present study, dS exceeded dN in all comparisons (Tab. III.3 and Tab. III.4), as would be expected for genes under purifying selection (Nei, 1987). However, the relative rate test (Tajima, 1993) showed statistically significant increase in evolution rates in some genes of common carp (Tab. III.5 and Tab. III.7) and goldfish (Tab. III.6 and Tab. III.8). In addition, the test if non-synonymous mutations occur in such a way as to change some amino acid property of interest to a greater extent than is expected under random substitution (neutral evolution) provided evidence even of positive Darwinian selection in some genes.

According to a survey by Endo et al. (1996), positive selection affecting entire protein-coding sequences is suspected in only very few cases. In their study of 3,595 groups of homologous sequences, they found only 17 gene groups (about 0.45%) in which the ratio of non-synonymous to synonymous substitutions was significantly larger than 1, i.e. in which positive selection may have played an important role in their evolution. Interestingly, 9 out of the 17 gene groups consisted of genes encoding surface antigens of parasites and viruses. If whole genes are used for analysis then positive selection affecting only parts of genes or individual sites might have gone undetected.

The growth hormone genes of common carp show increase in the evolution rate in one copy (*GH I*) (Tajima’s test). The p_{NR}/p_{NC} ratio is close to 1 for polarity, that means relaxation of functional constraints in one of the duplicates. Therefore, the activity of proteins from *GH* paralogues could be different and it should be tested. On the other hand, the expression could be time or tissue separated.

Many of the duplicated genes from the present study that do not show increased evolution rate (Tajima’s relative rate test) show relaxed functional constraints if p_{NR} and p_{NC} were compared.

The paralogues for *c - myc* of common carp don't demonstrate increased evolution rate, but the p_{NR}/p_{NC} value is 0.98 and near to 1 in this study. Futami et al. (2005) have tested the activity of *c - myc* paralogues of common carp and found that they have functional divergence.

Therefore, for analysis of evolution on DNA level of two duplicated genes several different methods to sift the natural selection should be used at the same time. Zhang (2003) states that the current statistical methods in detection of selection on the sequence analysis level are not yet enough efficient.

The different evolution rates between *GH II* and *c - myc II* orthologues of common carp and goldfish could indicate different evolution of these genes in these two fish species. Maybe mechanisms like this have conducted to the divergence of species that have the same tetraploid ancestor.

Increased evolution rates, relaxation of functional constraints or positive Darwinian selection were also found in some genes playing an important role in action or regulation of growth hormone.

An increased evolution rate was found for one of paralogues of the *IGF - I* gene in common carp (Tab. III.5 and Tab. III.7). IGF - I is an important mediator of the action of growth hormone during postnatal life in mammals and is synthesized predominantly in the liver (Froesch et al., 1985). The GH dependence of liver IGF - I expression has been observed in a number of teleost species. Injection of GH increased IGF - I mRNA and the serum level of the growth factor in salmonids, seabream and common carp (Cao et al., 1989; Shamblott et al., 1995; Vong et al., 2003). GH induction of IGF - I expression was also observed in primary hepatocytes of tilapia (Schmid et al., 2000). On the other hand, the secretion of GH is suppressed by IGF - I in human by actions at the level of GH gene transcription (Yamashita et al., 1986). The release of GH from teleost pituitary glands is similarly suppressed by IGF - I through mechanisms that are additive with SRIF (Perez-Sanchez, 1992). IGF - I-induced GH mRNA suppression may represent an autoregulatory component in GH regulation.

The GH receptor and *c - myc* genes of common carp (Tab. III.10) also show relaxation of functional constraints for polarity in one of their paralogues. The action of growth hormone involves binding with growth hormone receptor. Growth hormone receptors are present in fish in liver, kidney, skin, muscle, fat, spleen, brain, ovary, testis and gill (Gray et al., 1990; Hirano, 1991; Yao et al., 1991; Sakamoto & Hirano, 1991). The analysis of expression patterns of the two growth hormone receptor variants of common

carp and their ability to bind the GH could therefore be useful for understanding of growth mechanisms in this fish species.

The expression of *c - myc* is directly influenced by GH in liver (Murphy et al., 1987). The proto-oncogenes *c - myc* are thought to be one of the most important genes in controlling cell proliferation (Roy et al., 1993). They are crucial for the cell cycle, cell growth, differentiation, apoptosis, transformation, genomic instability and angiogenesis, and are highly conserved in vertebrates (Oster et al., 2002). As mentioned above, *c - myc* gene products show different activity in common carp (Futamy, 2005).

The *Pit-1* and *GRH* genes in common carp have p_{NR}/p_{NC} ratio >1 for change of polarity and in addition *Pit-1* genes of common carp show p_{NR}/p_{NC} ratio >1 for change of charge what is evidence for positive Darwinian selection in one of the paralogues.

Pituitary expression of GH is largely controlled by Pit-1 transcription factor, which has been located in normal and tumorous rat and human pituitaries and normal trout and common carp pituitaries (Yamada, 1993; Argenton et al., 1993; Delhase et al., 1993; Lloyd et al., 1993a,b; Kausel et al., 1999). Deficiency or dysfunction of Pit-1 in subsets of human and mouse (Snell) dwarfs is also associated with an absence or deficiency of GH biosynthesis (Pfaffle et al., 1992; Voss & Rosenfeld 1992; Li et al., 1990) whereas its ontogenetic appearance in the pituitary gland correlates with the activation of the GH gene (Dolle et al., 1990).

Therefore, the future analysis of activity and expression patterns of paralogous genes that are important for growth hormone action or regulation and that show differences in evolution rate between their paralogues or relaxation of functional constraints and especially positive Darwinian selection in one of their paralogues could be important for understanding of growth and development regulation in the tetraploid common carp.

Comparison of coding sequences of duplicate genes of the tetraploid frog *Xenopus laevis* showed no evidence for positive Darwinian selection or increase in evolution rate (Hughes & Hughes, 1993). The duplicated genes of *Xenopus laevis* were compared with orthologues of human in relative rate test. In contrast, in the present study on common carp and goldfish duplicated loci some paralogues could be found with increased evolution rate and/or relaxed functional constraints in one of them, some pairs of duplicated genes with evidence of positive Darwinian selection in one of the paralogues and at the same time some with evidence of purifying selection (Tab. III.5; Tab. III.6; Tab. III.7; Tab. III.8; Tab. III.9 and Tab. III.10).

Wallis et al. (1994, 1996) studied the evolution of mammalian growth hormone genes. They found that growth hormone genes evolve quite slowly throughout most

mammalian evolution. The mean evolutionary rate is approximately 0.3×10^{-9} replacements per amino acid site per year. There are, however, two independent bursts of rapid evolution, one prior to the divergence among ruminants, and one before primate divergence. Wallis et al. (1994) estimated that during these short times, which constitute less than 10% of the total evolutionary time, there was a 20-fold and a 40-fold increase in the rate of evolution in ruminants and primates, respectively. Wallis et al. (1996) computed the ratio of non-synonymous to ratios of synonymous substitutions during the slow and rapid phases of evolution. The results showed that during the rapid phases of evolution, there is a significant increase in the ratios of non-synonymous substitution/ratio of synonymous substitution values, indicative of either positive selection or relaxation of selection.

If that scenario is possible for other primate genes then application of human orthologues as sequences for analysis of evolution rate between duplicated genes in tetraploid fish or amphibian species in relative rate tests could conduce that differences between paralogues could be overlooked.

In the present study the common carp and goldfish paralogues were compared with zebrafish orthologues. All three fish species belong to the same cyprinid fish family. It may make sense to compare paralogues with orthologues from several species of different classes in relative rate tests and then to compare the results with each other. The difficulty is that not all sequences are available in GenBank.

For *POMC*, growth hormone receptor and prolactin receptor two cDNA sequences for two genes from common carp but only one cDNA sequence from goldfish were found. This could be explained by the incomplete nature of GenBank. On the other hand, one of the paralogues of these genes could be silenced in goldfish. Lynch & Conery (2000) proposed that the loss of different duplicates in geographically isolated populations could reduce the fecundity of hybrids. Genome duplication provides that all of the genes in the genome are duplicated. Divergent resolution of thousands of genes might be a very powerful isolating mechanism.

Consequently, a more precisely study how different duplicated genes of common carp and goldfish are evolving should be done. The evolution on DNA level of other paralogues of common carp and goldfish should be analysed. More detailed molecular and functional studies should be carried out. The expression patterns of duplicated genes and biological activity of their proteins should be tested. This could help to answer the question how whole genome duplication could conduce to the formation of new species and maybe if genome duplication could lead to complexity in vertebrates.

Chapter IV Expression of the two common carp growth hormone genes in relation to ontogenesis, sex and water temperature

IV.1 Introduction

IV.1.1 Growth hormone in growth, development, and maintenance of energy homeostasis

GH plays a role both in the growth and in development of teleosts (Donaldson et al., 1979; Higg et al., 1975; Degani & Gallagher, 1985). Strong evidence suggested that GH is an important regulator of gonadal growth (Fostier et al., 1983; Young et al., 1983; Singh et al., 1988; Van der Kraak et al., 1990; Le Gac et al., 1992; Singh & Thomas, 1993). In a study on the effect of GH on the stimulation of gonadotropin (GTH) and secretion of steroids it was suggested that GH has a direct modulatory effect on GTH-stimulated steroid production, and also that GH may be an important regulator of follicular development in the goldfish (Van der Kraak et al., 1990). In killifish, injection of recombinant salmon GH has stimulated the growth of the gonad and raised the level of sex steroids in the blood. Similarly, injection of recombinant rainbow trout GH raised the E₂ (estradiol) level in immature trout and killifish (Singh et al., 1988).

Bjornsson et al. (1994) found that the GH level in plasma of Atlantic salmon, *Salmo salar*, is higher during sexual maturity than in immature fish. In the white sucker, *Catostomus commersoni*, the GH level in blood of the female rises during ovulation and remains high after spawning (Stacey, 1984). Sumpter et al. (1991) demonstrated in rainbow trout that the GH level increases in plasma during ovulation in females and at the beginning of sperm production in males, it is concomitant with an increase in the level of 17 α 20 β – dihydroxy progesterone (Le Gac et al., 1992). In hypophysectomized eels, vitellogenesis does not rise in response to estradiol unless it is administered together with GH, which is also ineffective when administered on its own (Burzawa-Gerard et al., 1992). There are also a number of reports (Nashioka et al., 1988) showing that reproduction hormones affect the secretion of GH, e.g. E₂ and ethinylestradiol. In a study on the effects of testosterone (T) and E₂ on serum GH in female goldfish during the reproduction season a maximum GH level in spring and a minimum in summer were found, along with the fact that E₂ affected the GH level during the reproductive cycle while T did not (Trudeau et al., 1992). These authors suggested that gonadal steroids play a role in the control of GH secretion, since there is marked sexual dimorphism in secretory patterns of GH in teleosts.

Degani et al. (1996, 1998) found that *in vitro* secretion of GH from primary pituitary cell cultures of common carp increased under the influence of sex steroids. The extent of increase rose with the level of treatment with estradiol, testosterone, and 17 α -hydroxyprogesterone. The authors suggested that before maturation, when the steroid level is very low, the secretion of GH affects fish growth. The steroid levels rise during gonadal development, leading to increased secretion of GH. When steroids reach their maximum level, GH declines.

An increase in plasma GH is observed following a decrease of protein level in ration and dietary energy content (Perez-Sanchez et al., 1995; Company et al., 1999). A pronounced increase in circulating GH was also observed in gilthead sea bream fed diets with poorly balanced amino acid profiles (Gomez-Requeni et al., 2004). This observation was related to a decrease in plasma IGF - I levels in concurrence with a reduced expression of hepatic *IGF - I* and *GHR* genes, the features characteristic for catabolic state. Fox et al. (2006) found that in tilapia, *Oreochromis mossambicus*, fasting resulted in significant reductions in body weight and specific growth rate. Plasma GH and pituitary GH mRNA levels were significantly elevated, whereas significant reductions were observed in plasma IGF - I and hepatic IGF - I mRNA levels. There was a significant negative correlation between plasma levels of GH and IGF - I. Therefore, liver GH desensitization and the inverse relationships between plasma GH and IGF - I concentrations represent a conserved mechanism for the preferential utilization of mobilized substrates to maintain energy homeostasis rather than cell growth and proliferation under either reduced nutrition or malnutrition (reviewed in Renaville et al., 2002)

Isoosmotic salinity (12 ppt) adaptation resulted in increased GH transcript and protein in a comparative study of freshwater (0 ppt) and seawater (33 ppt) adapted sea bream. Sea bream acclimated to cold temperature (12°C) also had higher amounts of pituitary GH transcript and protein as compared to warm temperature (25°C) acclimated fish (Deane et al., 2006).

Therefore, the growth hormone is important for growth, sexual maturation, reproduction, catabolism and adaptation of fish to different environmental conditions. The common carp has two growth hormone genes and it is very useful to know which importance each of them has for the above described processes.

IV.1.2 Differences in the timing and/or pattern of expression of duplicated genes

Theoretical population genetics predicts that both duplicated genes can be stably maintained if they differ in some aspects of their functions (Nowak et al., 1997), which can

occur through subfunctionalization. In that case each daughter gene adopts part of the functions of their parental gene (Jensen, 1976; Orgel, 1977; Hughes, 1994). One potentially important form of subfunctionalization is division of gene expression after duplication (Force et al., 1999). The duplicated genes often differ in the timing and/or pattern of expression (Ferris & Whitt, 1979; Hughes & Hughes, 1993).

In rainbow trout (a fish species of tetraploid origin) differential levels of the mRNAs have been observed for the two *GH* genes *rtGHI* and *rtGH II* in relation to sex and developmental stage (age). The amounts of *rtGHI* products in 10 days old fry and 2 years old females were higher than those of *rtGH II*. Additionally, it has been found that the *rtGH II* gene does not have an estrogen response element which is, on the contrary, present in the *rtGHI* gene (Yang et al., 1997).

The duplicated *c - myc* genes of common carp have also shown some different expression pattern (Futami et al., 2001).

The timing and/or pattern of expression of the two growth hormone genes of common carp could also be diverse. To understand the roles of the two *GH* genes it should be examined whether these genes show similar or different levels of expression during ontogenesis, in males and females and under different environmental conditions.

IV.1.3 Advantage of the real-time PCR technique in examination of gene expression

Reverse transcription followed by PCR is the most powerful tool to amplify small amounts of mRNA (Wang et al., 1989). The rapid cycle PCR in the LightCycler system offers stringent reaction conditions to all PCR components and leads to a primer sensitive and template specific PCR because of its high ramping rates, limited annealing and elongation time (Wittwer et al., 1997). The application of fluorescence techniques to real-time PCR combines the PCR amplification, product detection and quantification of newly synthesized DNA.

The LightCycler HybProbe format is based on the principle of fluorescence resonance energy transfer (FRET). Two sequence-specific oligonucleotide probes are labeled with different dyes (donor and acceptor), and are added to the reaction mix in addition to the PCR primers. HybProbe probes hybridize to the target sequences on the amplified DNA fragment during the annealing phase in head-to-tail arrangement, thereby bringing the two dyes into close proximity. The donor dye (fluorescein) is excited by the blue light LED source. When the two dyes are close to each other, the energy emitted by the donor dye excites the acceptor dye attached to the second HybProbe probe, which then emits

fluorescent light at a different wavelength (Fig. IV.1). The amount of fluorescence is directly proportional to the amount of target DNA generated during the PCR process. HybProbe probes are displaced during the elongation and denaturation steps. In a HybProbe analysis, the presence of a specific amplification product is quantitatively recorded by an increase in fluorescence. No fluorescence is recorded in the absence of specific template.

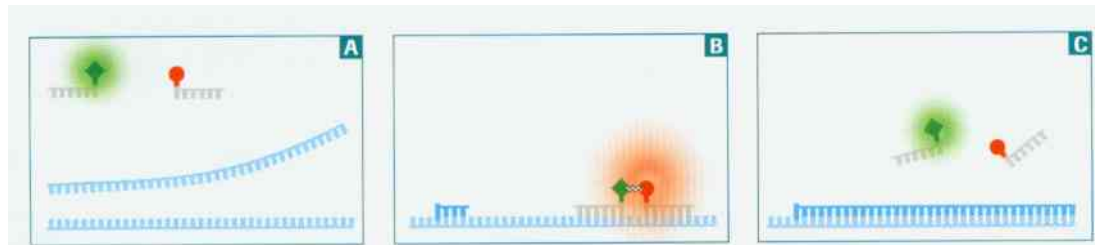


Figure IV.1 Schematic presentation of the HybProbe format (from information brochure “The LightCycler System” of Roche Applied Science). A. The donor-dye probe is labeled with fluorescein at the 3’ end and the acceptor-dye probe is labeled with LightCycler Red at the 5’ end. Hybridization does not take place during the denaturation phase of PCR and, thus the distance between the dyes is too large to allow energy transfer to occur. B. During the annealing phase, the probes hybridize to the amplified DNA fragment in a close head-to tail arrangement. When fluorescein is excited by the light from the LED, it emits green fluorescent light, transferring the energy to LightCycler Red, which then emits red fluorescent light. This red fluorescence is measured at the end of each annealing step, when the fluorescence intensity is highest. C. After annealing, the temperature is raised and the HybProbe probe is displaced during elongation. At the end of this step, the PCR product is double-stranded and the displaced HybProbe probes are again too far apart to allow FRET to occur.

The real-time PCR detection makes a kinetic quantification possible. This technique allows data analysis in the phase of constant amplification efficiency (log-linear phase) (Orlando et al., 1998). PCR amplification is template concentration dependent, but reactions with low starting copy number can reach the same plateau as reactions that started with higher template concentrations and/or a different PCR efficiency. In contrast to analysis during end-point-PCR, analysis in the log-linear phase produces data that are much more accurate (Fig. IV.2). Since amplification efficiency is constant, the amount of starting material can be determined very precisely. It is necessary to determine the crossing points (Cp) for the transcript. Cp is defined as the point at which the fluorescence rises appreciably above the background fluorescence (the cycle number of PCR at detection threshold).

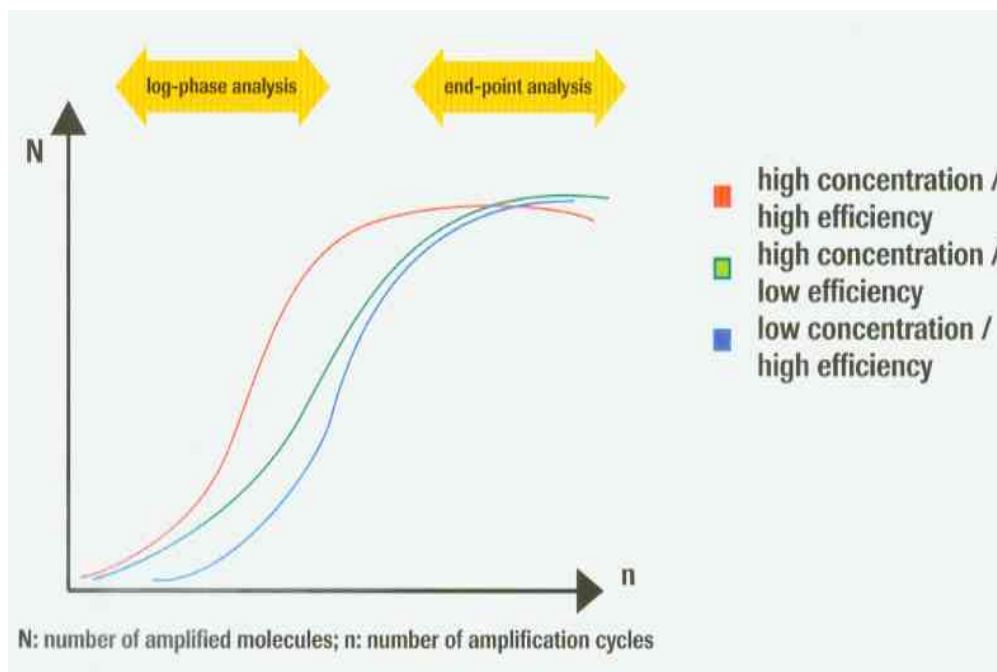


Figure IV.2 PCR and the problem of quantification (from Technical Note N LC 13/2001 “Relative Quantification” of Roche Applied Science).

Numerous studies have already shown that housekeeping genes are regulated and may vary under experimental conditions (Zhang & Snyder, 1992; Bhatia et al., 1994; Bereta & Bereta, 1995; Chang et al., 1998; Sarmiento et al., 2000). The common carp is tetraploid and has a lot of genes in duplicate and both copies could be expressed. Duplicated genes could have very similar sequences but different expression patterns. This makes it more difficult to find a housekeeping gene in this fish species. It is especially troubled if expression of the two growth hormone genes of common carp needs to be analysed during ontogenesis, in different sex and environmental conditions. Through subfunctionalization duplicated housekeeping genes could also be differently regulated during all of the above listed stages and conditions.

Normally, the procedure of the target transcript normalization to the housekeeping gene needs production of the external standard materials of both the target transcript and the housekeeping gene. In addition, optimization and validation of the calibration curves have to be done. A simpler approach can be used. Such model calculates the relative expression ratio only from the real-time PCR efficiencies (E) and crossing point (Cp) deviations of the investigated transcripts and stable constant reference (Pfaffl, 2001).

This model could be further simplified when studying possible differences in the expression of the two growth hormone genes of common carp during ontogenesis, sexual maturation and different environmental conditions. As soon as only the differences in mRNA levels between *GH I* and *GH II* genes in each stage and condition are in question,

the use of a housekeeping gene can be avoided. The calculation of the *GH II/GH I* mRNA level ratio could be based on the crossing points of *GH I* and *GH II* cDNA transcripts in the same sample and the efficiency of the PCR of *GH I* and *GH II*. This approach allowed circumventing the problem to find a housekeeping gene in this tetraploid fish species which amount is constant during ontogenesis and under different environmental conditions. The calculations are made according to the general PCR equation:

$$N_{GH I} = N_{GH I 0} \times E_{GH I}^{Cp_{GH I}},$$

$$N_{GH II} = N_{GH II 0} \times E_{GH II}^{Cp_{GH II}},$$

where:

$N_{GH I 0}$ and $N_{GH II 0}$ are initial numbers of *GH I* and *GH II* cDNA molecules. $N_{GH I}$ and $N_{GH II}$ are numbers of *GH I* and *GH II* cDNA molecules at detection threshold Cp. $E_{GH I}$ and $E_{GH II}$ are efficiencies of *GH I* and *GH II* cDNA amplification. $Cp_{GH I}$ and $Cp_{GH II}$ are cycle numbers at *GH I* and *GH II* cDNA detection threshold (crossing point). The *GH II/GH I* ratio is calculated as follows:

$$N_{GH II} / N_{GH I} = N_{GH II 0} \times E_{GH II}^{Cp_{GH II}} / N_{GH I 0} \times E_{GH I}^{Cp_{GH I}}$$

The amplicon numbers of *GH I* and *GH II* cDNA become identical at a certain fluorescence detection level (detection threshold).

$$N_{GH II} / N_{GH I} = 1, \text{ and}$$

$$N_{GH II 0} \times E_{GH II}^{Cp_{GH II}} / N_{GH I 0} \times E_{GH I}^{Cp_{GH I}} = 1$$

and it could be obtained

$$N_{GH II 0} / N_{GH I 0} = E_{GH I}^{Cp_{GH I}} / E_{GH II}^{Cp_{GH II}}$$

The last equation demonstrates that the ratio of the initial number of *GH II/GH I* cDNA molecules is equal to the ratio $E_{GH I}^{Cp_{GH I}} / E_{GH II}^{Cp_{GH II}}$, that used only PCR efficiencies of *GH I* and *GH II* cDNA templates and crossing points of both transcripts.

The question of the present study is if there are differences in levels of *GH I* and *GH II* mRNA in common carp during ontogenesis, in sex and under different environmental conditions. Therefore, the ratio of *GH II/GH I* mRNA level is a reasonable indicator. If the ratio will be 1, then both genes are transcribed with the same intensity and if it will be less or greater than 1, then there are differences in transcription of the two growth hormone genes of common carp.

The amplification of both genes could be carried out in one glass capillary with the same primers but specific for each gene cDNA HybProbe probes emitting fluorescent light at a different wavelength. This approach reduces the impact of pipetting errors, ensuring very precise relative quantification measurements.

This model also allows to standardize each reaction run with respect to RNA integrity, RT efficiency or cDNA sample loading variation. The reproducibility of the RT step varies greatly between tissues, the applied RNA isolation methodology (Mannhalter et al., 2000) and the RT enzymes used (Wong et al., 1998). During two-step RT-PCR using poly-T primers all possible interferences will influence both transcripts of *GH I* as well as *GH II* in parallel. Occurring background interferences retrieved from extracted tissue components, like enzyme inhibitors, and cDNA synthesis efficiency will be related to *GH I* and *GH II* similarly. Any source of error during RT will be compensated through the model itself.

IV.2 Material and Methods

IV.2.1 Experimental fish

Mirror carp fry were obtained from Fischzucht Jänschwalde GmbH fish farm. The fish were maintained at natural photoperiods and at 23-24°C, fry in aquaria and larger fish in tanks. Carp fry was fed with live *Artemia nauplii* and older fish with commercial dry diet (AQUAVALENT; Märkische Kraftfutter GmbH). For adaptation to cold temperature (hibernation) part of fish were transferred to outdoor fishponds from October to February without feeding. These fish were collected at a water temperature of 2°C at the end of winter.

IV.2.2 Sample collection and tissue RNA preparation

Common carp fry were collected at 15, 22, 29 days after hatching (30 fry per tube, six tubes per stage); at 47 and 63 days after hatching (two or three heads per tube, six tubes per stage); 85 days after hatching (three whole brains per tube, six tubes per stage). One

pituitary gland per tube was collected from 1 year old fish (10 tubes in total), 3 years old males (9 tubes in total), and 3 years old females (10 tubes in total). In addition, pituitary glands were collected from 10 months old fish that were cold adapted (one pituitary gland per tube, total of 20 tubes).

In case of the cold adapted fish, no cDNA could be obtained after RNA isolation from single pituitary glands and subsequent reverse transcription and PCR. Even pooling of two or three pituitary glands before RNA isolation was without success. Only after pooling of 10 single pituitary glands sufficient amounts of RNA could be isolated and a PCR product was obtained.

All samples were put in TRIzol reagent (Invitrogen), frozen in liquid nitrogen and kept at -80°C until RNA isolation. Total RNA was isolated from the fish tissues using TRIzol reagent (Invitrogen). Samples were homogenized with Ultra-Turrax (IKA Labortechnik). Chloroform was added, the mixture was vortexed vigorously, left for 5 min at room temperature and centrifuged for 15 min at 12,000 g, 4°C. RNA was extracted from the aqueous phase with isopropanol and precipitated by centrifugation (15 min, 12,000 g, 4°C). The pellet was washed with ethanol (75%, v/v), centrifuged (10 min, 12,000 g, 4°C), dried and resuspended in diethylpyrocarbonate (DEPC)-treated water. RNA samples were quantified by determining their absorbance at 260 and 280 nm with a spectrophotometer.

IV.2.3 First strand cDNA synthesis

Before cDNA synthesis all samples were treated with RNase-free Deoxyribonuclease I (Fermentas). Then, RNA (1 µg) was reverse transcribed in a 20 µl reaction volume for 5 min at 37°C, 60 min at 42°C and 10 min at 70°C. The reaction mixture consisted of 5 pmol oligo dT primer, 200 U RevertAid™ H Minus M-MuLV Reverse Transcriptase (Fermentas), in 1 x reverse transcriptase buffer, 10 mM dNTPs and 10 U RNase inhibitor (Fermentas).

IV.2.4 Quantitative RT-PCR and data analysis

Samples were analyzed by real-time PCR after cDNA synthesis using a LightCycler Instrument 1.5. The measurements for each sample were independently repeated three times. The amplification of PCR products was verified using agarose gel electrophoresis. For amplification and detection of *GHI* and *GHI* cDNA a degenerative primer pair was constructed:

forward: 5' -GTGCTCATCMAGGSATGTCTCG-3' ,

reverse 5' -AGRTAAKGTTTAAATATRGGAGGG-3' .

The hybridization probes for *GHI* were:

5' -CCGGTGTATTGTTAGTCAATGCCT-fluorescein-3'

5' -LC Red 640-TAACACATTTGTGCTTTGCTGCA-3'

The hybridization probes for *GH II* were:

5' -AGAGGGTGCCTTTGTATTGCTAGTCAGT-fluorescein-3'

5' -LC Red 705-CCTGCAAATCTAAGACTAGTTTAAGTCCTT-3'

The primers and hybridization probes were generated by Tib-Molbiol (Berlin, Germany) and lie in non-coding 3' ends of *GHI* and *GH II* cDNA. The cDNA sequences of *GHI* and *GH II* (Koren et al., 1989; Chao et al., 1989) and 3' end of *GHI* and *GH II* gene sequences of common carp obtained in the second part of the present study were used for construction of primers and hybridization probes.

The HybProbe probe for *GHI* cDNA had an acceptor with LightCycler Red 640 dye emitting fluorescent light at 640 nm and for *GH II* cDNA the acceptor in HybProbe probe was labeled with LightCycler Red 705 dye emitting fluorescent light at 705 nm. The different emission wavelengths of these fluorescent dyes can be measured in separate fluorescence channels of the LightCycler Instrument (dual-color experiments). This made it possible to analyse the expression of *GHI* and *GH II* in one PCR reaction tube. Cross-talk between fluorescence channels was corrected using a color-compensation file. No fluorescent signal was obtained using only *GHI* DNA for real-time PCR with HybProbe probe for *GH II* and vice versa (that shows that HybProbe probe for growth hormone *GHI* cDNA is specific only to *GHI* and the other only to *GH II*).

Reactions were prepared in 10 µl in 1.5 mm high quality borosilicate glass LightCycler Capillaries (Roche) containing the following components: 0.125 µl LightCycler Uracil-DNA Glycosylase; 2 µl of the LightCycler FastStart DNA Master^{PLUS} HybProbe, 5 x conc. (contains FastStart Taq DNA Polymerase, reaction buffer, MgCl₂, and dNTP (with dUTP instead of dTTP)) (Roche); 1 µl of a mixture of forward and reverse primers (2 µM each) and HybProbe probes (2 µM each); 4.375 µl nuclease-free water; 2.5 µl cDNA template. The 0.25 U glycosylase were added to prevent carry-over contamination.

Standard operation procedure for the LightCycler 1.5 instrument was followed in this study. This included the use of all default program settings according to the LightCycler Operator's Manual Version 3.5. Amplification was performed under the following conditions: 10 min at 40°C, 10 min pre-incubation step at 95°C, followed by 45 cycles of 10 sec at 95°C, 10 sec at 56°C, and 5 sec at 72°C (temperature ramp 20°C/sec).

Cp determinations were performed for each reaction using Second Derivative Maximum Method.

The data were first analysed with LightCycler Software Version 3.5 and then the relative quantification was carried out using LightCycler Relative Quantification Software Version 1.

The PCR efficiencies of cDNA for *GHI* and *GHII* in one-capillary were analyzed with LightCycler Relative Quantification Software Version 1 and were used for comparing of the expression of *GHII* to *GHI*. A serial dilution of calibrator sample from 1 to 10⁻³ was prepared and obtained Cp values were plotted against log of dilution to get the efficiencies curve.

The statistical data analysis was aimed at establishing the variability range of the *GHII/GHI* mRNA level ratio and its dependence on ontogenesis, sexual maturation and environmental conditions. The analysis was performed using well-developed statistical methods including analysis of variance (*one-way ANOVA*) and hypothesis testing. Since most methods of the hypothesis testing rely on the assumption of normally distributed data, the normality test was performed for all samples. The *Lilliefors test*, an adaptation of the Kolmogorov-Smirnov test especially designed for short data series was used for normality testing. All statistical calculations were performed using procedures from the MathWorks Matlab Statistics Toolbox (www.mathworks.com).

IV.3 Results

The primary data of the *GH II/GH I* mRNA level ratios obtained for each sample are presented in Table A.3 in the Appendix together with means and standard deviations for each sample. The same table also presents the number of samples (6 to 10) analysed for each developmental stage. Table IV.1 summarizes means and standard deviations for each group.

Table IV-1 Means and standard deviations of *GH II/GH I* mRNA level ratios for each studied group.

GH II/GH I mRNA level ratios			
Group (age or sex)	Number of samples	Mean	STD
15 days	6	0.92	0.16
22 days	6	0.93	0.16
29 days	6	0.89	0.10
47 days	6	0.90	0.17
63 days	6	0.92	0.16
85 days	6	0.80	0.24
1 year old	10	0.77	0.18
3 years old female	10	0.72	0.16
3 years old male	9	0.75	0.24
Cold adapted 10 months old fish	10 pooled	0.31	

Each observation in the data sample is represented by a triplicate of values. Estimation of the variance inside these triplicates was performed in order to find out whether each triplet of values has a common mean. Straight application of statistical hypothesis testing to such short sets of values is rather difficult, but indirect verification can be achieved from the ANOVA methods comparing the variance inside triplicates against the variance between triplicates of samples from the same developmental stage or sex. If the variance between three values of triplicates for each sample is really small compared to the

variance between samples, one can consider them as representing the same value, which is chosen as the mean of the corresponding triplet. The results are presented in Table IV.2.

Table IV-2 One-way ANOVA statistics for each sample.

Sample	F statistics	p-value
'15 Days Old'	15.9472	6.1470e-005
'22 Days Old'	10.1227	5.5531e-004
'29 Days Old'	8.9733	9.6022e-004
'47 Days Old'	26.3420	4.4466e-006
'63 Days Old'	24.7202	6.2570e-006
'85 Days Old'	120.5846	7.8959e-010
'1 Year Old'	66.1050	6.5226e-013
'3 Years Old, Male'	74.5747	2.0506e-013
'3 Years Old, Female'	74.2664	3.3369e-012

The p-values are small for each developmental stage or sex, allowing us to assume that variances inside every observation triplet are negligible. Thus, the mean of the three readings could be accepted as a reliable value for further analysis. It should be noted that p-values in Table IV.2 are extremely small for fish of 85 days old and older, whereas the variance between samples is larger than that in younger fish (see Tab. A.3 in Appendix). This significant increase in variance during the fish growth indicates less reliability of the subsequent statistical analysis based on short data sets applied to adult fish.

Assuming negligibility of variations inside observation triplicates, every age group is presented in further analysis by a vector of triplicate means. Plots of empirical probability density functions (PDF) are presented in Figure IV.3 for each age group. Based on empirical PDF it is difficult to decide if the distribution is close to the normal one or not. The obvious reason for it is the short length of the data samples.

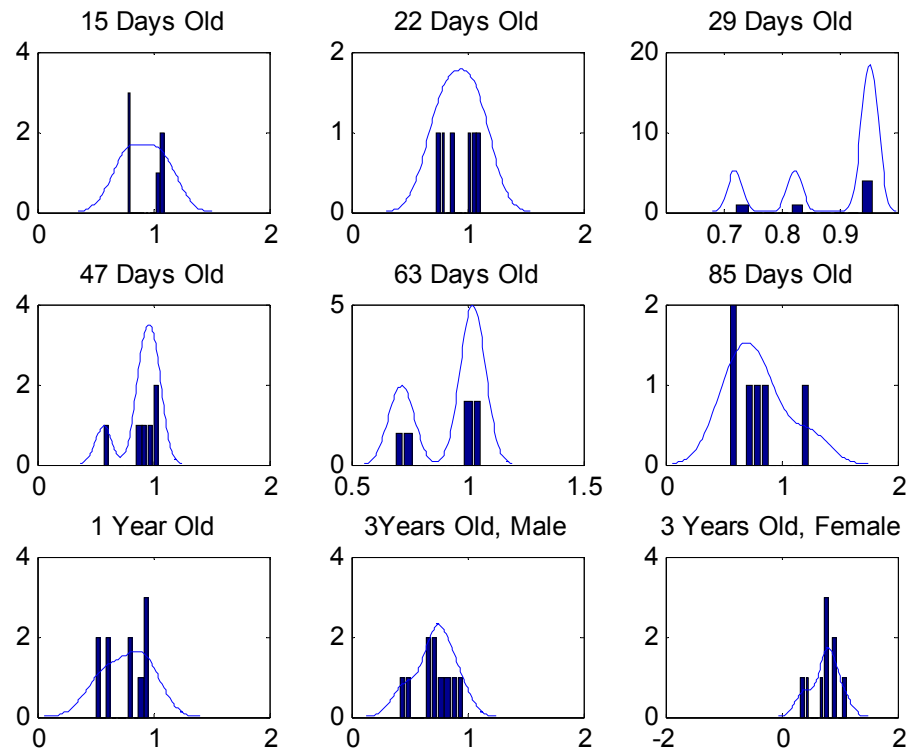


Figure IV.3 Empirical probability density functions of *GH II/GH I* mRNA level ratios for different age and sex groups.

Therefore, an alternative method was applied – Lilliefors test, a modification of the Kolmogorov-Smirnov test, proven to be efficient for short data series. The null-hypothesis for the test is that the samples are normally distributed. The logical 0 returned by each test indicates a failure at 95% level to reject the null-hypothesis. This failure may reflect normality in the population, but also it may reflect a lack of strong evidence against the null-hypothesis due to the small sample size. The results of Lilliefors test for all studied ontogenetic and sex groups are presented in Table IV.3.

Table IV-3 Results of Lilliefors statistics. Null-hypothesis – the samples are normally distributed in each stage. 0 - indicates a failure at 95% level to reject the null-hypothesis. 1 - indicates that null-hypothesis can be rejected.

Sample	Lilliefors test
'15 Days Old'	0
'22 Days Old'	0
'29 Days Old'	1
'47 Days Old'	0
'63 Days Old'	0
'85 Days Old'	0
'1 Year Old'	0
'3 Years Old, Male'	0
'3 Years Old, Female'	0

The normality hypothesis cannot be rejected for all samples except one (29 days old fish). This also means that usual hypothesis testing methods generally cannot be applied to the latter data set.

Additional information on the statistical characteristics of various age groups can be gained from the Box-Whiskers plot in Figure IV.4. The decrease of the median value and increase of the variance with age is clearly seen in the figure. Some samples reveal also significant asymmetry. Most apparently, that was the reason for the normality hypothesis rejection of the 29 days old fish.

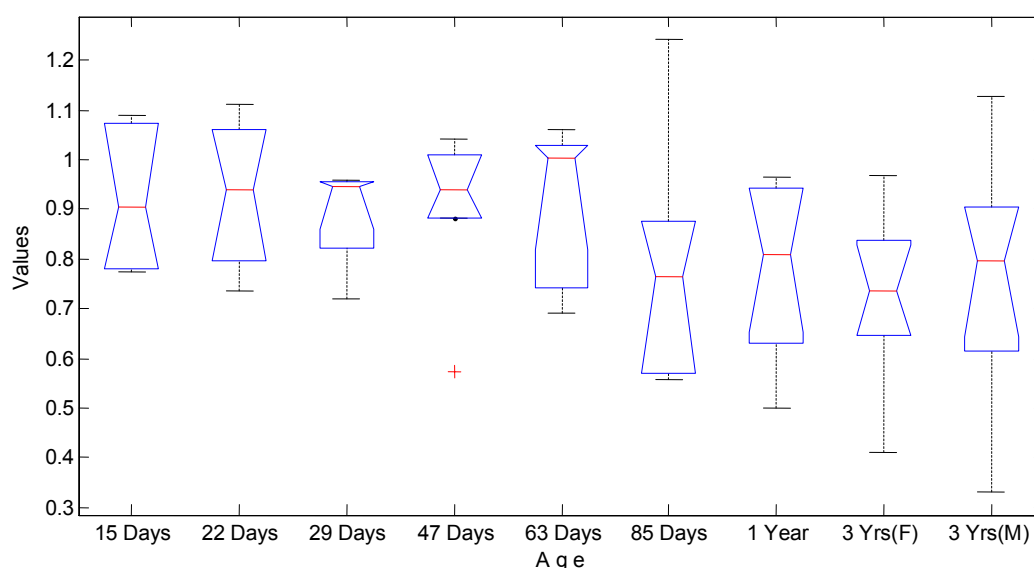


Figure IV.4 Box-Whiskers plot of *GH II/GH I* mRNA level ratios for different age and sex groups. Red lines are medians, boxes cover 50% of data in each sample, whiskers designate upper and lower quartiles, red crosses are possible outtakes.

In order to estimate the significance of the age-dependent decrease of the *GH II/GH I* mRNA level ratio, the hypothesis that the ratio is equal to 1 was tested for all ontogenetic and sex groups. The Welch t-statistics for the unknown variance was used. The null-hypothesis that the mean value is equal to 1 cannot be rejected for common carp younger than 1 year with 95% confidence level (Tab. IV.4). This provides evidence that both growth hormone genes have the same mRNA level in young fish. The null-hypothesis is rejected and the alternative hypothesis (mean value <1) is true for the adult fish (1 year and older). This means that mRNA level of *GH I* is higher than mRNA level of *GH II* in older fish. In addition, the hypothesis is rejected for the 29 days old fish, but applicability of the t-statistics to this age group is questionable since data in this group are not normally distributed (see results of the Lilliefors test above).

Table IV-4 Welch t-statistics for the unknown variance to test if the *GH II/GH I* mRNA level ratio is equal to one. Null-hypothesis – *GH II/GH I* ratio equal to one. 0 - indicates a failure at 95% level to reject the null-hypothesis. 1 - indicates that null-hypothesis can be rejected.

Sample	Welch t-statistics
15 Days Old	0
22 Days Old	0
29 Days Old	1
47 Days Old	0
63 Days Old	0
85 Days Old	0
1 Year Old	1
3 Years Old, Male	1
3 Years Old, Female	1

The plot of mean values presented in Figure IV.5 complements the results of the t-test. The mean ratio of mRNA levels for 1 year old and older fish range about the same value of 0.75 and are apparently different from the juvenile fish mean ratio of mRNA levels, which are all close to 0.9-1.0. The 85 days old sample demonstrates an intermediate value of 0.8 suggesting that the aging effect already starts to reveal itself on this stage. Additionally, the single value from the 10 months old cold-adapted fish is presented in Figure IV.5 demonstrating a significant difference in *GH I* and *GH II* mRNA levels.

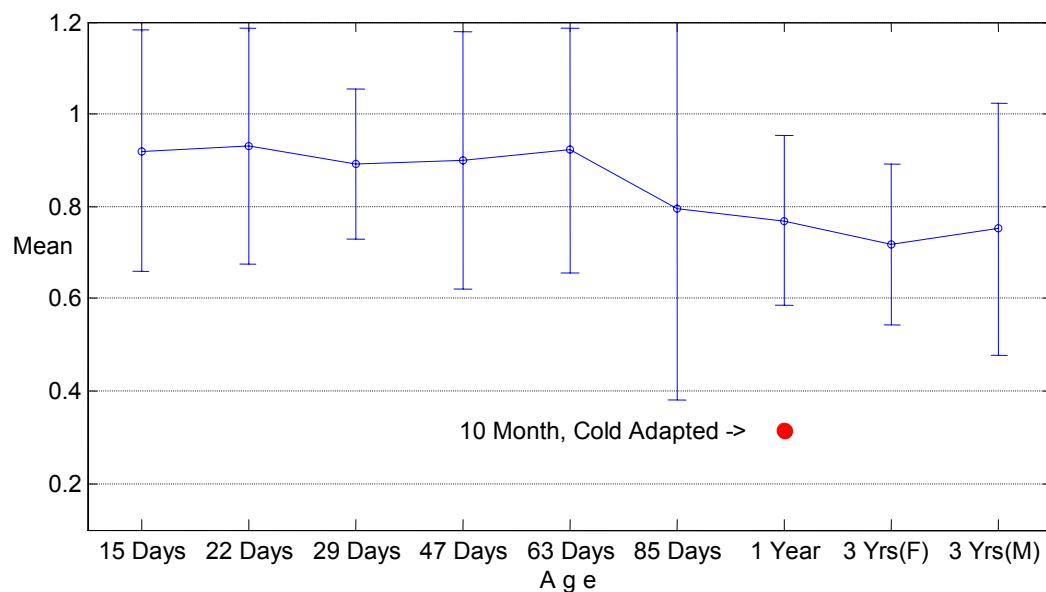


Figure IV.5 Mean values and 95% confidence intervals of *GH II/GH I* mRNA level ratios for different age and sex groups. The red spot represented the single value of *GH II/GH I* mRNA level from 10 months old cold-adapted fish.

IV.4 Discussion

The present study using real-time RT-PCR provides evidence for the active transcription of mRNA encoding for *GH I* and *GH II* in common carp fry and in pituitary glands of adult fish.

The levels of mRNA encoding for *GH I* and *GH II* were equivalent in all common carp fry groups except the 29 days old fish. But this group has not shown a normal distribution, the obvious reason for it is the short length of the data samples, and this result should therefore be excluded.

Growth in fish can be stimulated by the administration of GH (Degani & Gallagher, 1993; Degani et al., 1996). The equal levels of *GH I* and *GH II* mRNA in juvenile fish demonstrate that both genes are important for growth and development of common carp fry. It should be noted, however, that to determine the growth hormone mRNA level ratios not only pituitary glands but the whole fry (all somatic tissue) was used in the present study for stages 15 and 22 days after hatching; whole heads for stages 47 and 63 days and whole brain for stage 85 days. In fish, birds and mammals, the gene encoding for growth hormone is expressed not only in the pituitary gland but in many other tissues (brain, lymphocytes, placenta, mammary tissue, pineal gland, spermatogonia, primary spermatocytes, oocytes and granulose cells), and GH appear to regulate a host of cellular events through paracrine or

autocrine interactions (Izadyar et al., 1999; Butler & Le Roith, 2001; Harvey et al., 2004; Yada et al., 2005). As reported by several investigators (Funkenstein & Cohen, 1996; Yang et al., 1999; de Jesus et al., 2002), the presence of GH or mRNA encoding for these hormones in fish embryos prior to the emergence of a functional pituitary gland suggests that the hormones are not produced by elements of only the somatotrophic axis. Therefore, for better understanding of the function of the two growth hormone genes in common carp the *GH II/GH I* mRNA level ratios should also be measured in other tissues and in embryos.

In one year old fish and 3 years old fish of different sex *GH I* has a higher mRNA level than *GH II*. The ratio of *GH II/GH I* mRNA levels is significantly less than one. The gonadosomatic index was not measured in this study but the adult fish had developed gonads. The gonadal development has association with the production of growth hormone in pituitary and possibly affects the *GH II/GH I* ratio of mRNA levels in pituitary of common carp.

There is strong evidence that GH is an important regulator of gonadal growth and development in teleosts (Fostier et al., 1983; Higg et al., 1975; Young et al., 1983). Degani (1998) found that high concentrations (10^{-6} M) of the steroids E2, T and 17-P increase the release of GH from the carp pituitary *in vitro*. Similar results were obtained by Huggard & Habibi (1996) in an *in vitro* study, in which an increased mRNA level for GH in pituitary was found after treatment with 20 ng ml⁻¹ T. Comparing the *in vivo* and *in vitro* results, these authors suggested that T had a direct stimulatory effect on the GH mRNA level. Therefore, the more intensive production of mRNA for *GH I* than for *GH II* in one year old and older fish of different sex could serve as an evidence of subfunctionalization of *GH I* and *GH II* genes for sexual maturation and reproduction in common carp.

Degani (1996, 1998) analysed the pattern of GH synthesis in the pituitary and its secretion into the plasma of common carp during growth and sexual maturation, and concluded that before maturation the GH mainly affects growth. In mature females and males, the GH function is more concerned with reproduction. The present study demonstrates that during the growth both growth hormone genes have the same mRNA levels (the ratio of *GH II/GH I* mRNA levels is equal to one) and this is a sign that both of them should be important for growth and development of young fish. On the other hand, during oogenesis, spermatogenesis and gonadal development the ratio of *GH II/GH I* mRNA levels is significantly less than one and this could be a sign of different importance of the two growth hormone genes for sexual maturation and gonadal development.

The significant increase in variance of *GH II/GH I* mRNA level ratios observed in older fish groups reared at 23-24°C could be explained by asynchronous gonadal

development in these fish. Under temperate climate conditions, common carp spawn annually in response to an interaction of water temperature and light (day length). If one of these factors is kept constant, adult carp may finally cease cyclic gonadal maturation (Davies & Hanyu, 1986). As reported by Steffens (1980), carp permanently reared in warm water can be artificially reproduced out of season and even several times per year.

Alternatively, the lower variance of *GH II*/*GH I* mRNA level ratios observed in juvenile carp in comparison to older fish could also be an artefact resulting from differences in sample collection. In contrast to one year old and older common carp where measurements are based on single pituitary glands, pools of 30 fry or 2-3 heads/brains were analysed in juvenile carp. Since these pools already represent “mean” values of several individuals, their variance should be lower than that calculated from single individual measurements.

One single value from 10 months old cold-adapted fish presented in Figure IV.5 demonstrates a strong difference in *GH I* and *GH II* mRNA level. The expression of both growth hormone genes in common carp was possibly very low during hibernation and it was not possible to obtain PCR products if RNA isolation was done from only one, two or three pituitary glands. It was not possible to carry out a statistic test if this value is different from one or not, because it was only one mean. However, in this sample 10 pituitary glands from cold-adapted fish were pooled. Therefore, the value presented on Figure IV.5 could be taken into account. The strong deviation between *GH I* and *GH II* mRNA levels might indicate a diverse functional role of the two growth hormones of common carp for catabolic state. Exogenous GH administration has been shown to stimulate the intestinal amino acid transport and to increase intestinal mass in some teleosts (Collie & Stevens, 1985; Sun & Farmanfarmaian, 1992), suggesting that this may be one mechanism by which GH increases food conversion efficiency. The *GH I* of common carp is more intensively expressed under reduced nutrition or malnutrition and cold environment. In such conditions growth hormone plays an important role for preferential utilization of mobilized substrate to maintain energy homeostasis, rather than cell growth and proliferation. During this period the expression of the *IGF - I* gene in liver and *GHR* gene is strongly reduced (Perez-Sanchez et al., 1995; Company et al., 1999; Gomez-Requeni et al., 2004; Fox et al., 2006; Renaville et al., 2002).

As theoretical population genetics predicts, both duplicated genes can be maintained when they differ in some aspects of their functions (Nowak et al., 1997), which can occur by subfunctionalization (Jensen, 1976; Orgel, 1977; Hughes, 1994). Lynch & Force (2000) formulated a process for differentiation of regulatory regions of duplicated genes based on their subfunctionalization. In their model, loss of function of one of the duplicate genes at a

certain developmental time or space alters subsequent subfunctionalization which ultimately prevents coding region deterioration. As compared with differentiation of protein coding regions, initial differentiation of regulatory regions seems to depend more on drift rather than on selection (Lynch & Force, 2000; Zuckerkandl, 2001).

The present study shows that the two growth hormone genes of common carp subfunctionalized in division of their gene expression during sexual maturation and as a reaction to environmental conditions such as low temperature and starvation. It is possible that differentiation of regulatory elements occurred during evolution of the two growth hormone genes of common carp.

The differences between the two growth hormone genes of common carp in the first intron as described in the second chapter of the present study are very large. There are four different microsatellite sites in the first intron of *GH II*. They were not found or were very short in the first intron of *GH I*: in the first intron of *GH I* only very short (TA)₂ is present. In the first intron of *GH II* (TA) was repeated from 6 to 32 times in different alleles (Alignments A.1 and A.2; Tab. A.1 and A.2 in Appendix). On the basis of multiple alignments of the *GH II* gene alleles, a hairpin motif was predicted formed by microsatellites TA in the first intron. It has been demonstrated that the first intron of several genes such as $\alpha 1(I)$ collagen of human, human platelet derived growth factor-B, human Interferon- γ and human epidermal growth factor receptor has an important regulatory function (Bornstein et al., 1988; Franklin et al., 1991; Sica et al., 1992; Chrysogelos, 1993). On the other hand, secondary structures in introns could influence gene transcription or mRNA processing (Kirby et al., 1995; Buratti & Baralle, 2004). It should be tested in future if the differential regulation of the two growth hormone genes of common carp is possibly due to differences in sequences, differences in secondary structures or differences in DNA curvature in the first intron of the genes. On the other hand, promoter regions of both genes should be sequenced and analysed in the context of the different regulation of the two growth hormone genes.

The concept of the developed online PCR assay could also be useful for fast analysis of expression sharing between other duplicated genes (especially in tetraploid species). This approach allows analysing the expression rate of two duplicated genes without using a housekeeping gene that could be very difficult to find in species of tetraploid origin.

Chapter V Elaboration of a simple, fast and highly productive approach for the production of soluble common carp recombinant growth hormone

V.1 Introduction

V.1.1 Studies of functional divergence of duplicated genes

There are six amino acid substitutions between mature proteins of the two duplicated common carp growth hormone genes (Alignment 3 in Appendix). The genes show increase in the evolution rate in one copy: *GHI* (Tajima's test). The p_{NR}/p_{NC} ratio is close to 1 for polarity between *GHI* and *GHII* (see Chapter III of this study), that could reflect the relaxation of functional constraints in one of the paralogues. Therefore, the activity of proteins from *GH* paralogues of common carp could be different and should be tested *in vitro* or *in vivo*.

The studies described below demonstrate that an investigation of the activity of the recombinant proteins of duplicated genes is important for understanding the evolution and functional role of both duplicates.

Futami et al. (2005) studied the functional divergence of duplicated *c - myc* (*CAM1* and *CAM2*) genes in common carp. The proto-oncogene *c - myc* is thought to be one of the most important genes in controlling cell proliferation (Roy et al., 1993). It is crucial for the cell cycle, cell growth, differentiation, apoptosis, transformation, genomic instability and angiogenesis, and is highly conserved in vertebrates (Oster et al., 2002). The proteins of the Myc family form heterodimers with the product of *Max* gene, the Myc/Max heterodimer can function as a transcription activator (Amati et al., 1992; Kato et al., 1992). The affinity between the Max protein and the two *c - myc* proteins of common carp translated *in vitro* showed that CAM2 binds to Max more tightly than does CAM1, which may explain the lower transcriptional activity of genes targeted by CAM1 (Futami et al., 2005).

Zhang et al. (2002) have shown how a duplicated copy of a gene encoding a pancreatic enzyme has evolved. They were particularly interested in a pancreatic enzyme, RNASE 1, which breaks down bacterial RNA. Most primates have one gene encoding the enzyme, but the researchers found that the douc langur, a colobine monkey from Asia, has two: one encodes RNASE 1, and its duplicate encodes a new enzyme, which they dubbed RNASE 1B. Through a series of computations and experiments, the researchers determined that the original gene encoding RNASE 1 remained unchanged after duplication, but its

twin, which encodes RNASE 1B, changed rapidly. They therefore hypothesized that the substitutions may have changed RNASE 1B in catalyzing the digestion of RNA. To test this hypothesis, they prepared recombinant proteins from douc langur RNASE 1B as well as RNASE 1 genes of human, rhesus monkey and douc langur, and examined their ribonucleolytic activities at different pH levels in standard RNase assay against yeast tRNA. It was determined that the optimal pH for human RNASE 1 is 7.4, a value that is within the pH range (7.4-8.0) measured in the small intestine of humans. The same optimal pH was observed for RNASE 1 of rhesus monkey and douc langur. Probably because of foregut fermentation and related changes in digestive physiology, the pH in the small intestine of colobine monkeys shifts to 6-7. Notably, the optimal pH for douc langur RNASE 1B was found to be 6.3. At pH 6.3, RNASE 1B is six times as active as RNASE 1 in digesting RNA, and the difference in their activities is statistically significant.

Sequence conservation of douc langur RNASE 1 after gene duplication and its unchanged optimal catalytic pH at 7.4 suggest that this protein acts in non-digestive processes. Of note, human RNASE 1 is found in many other tissues besides the pancreas (Futami et al., 1997) and has enzyme activity in degrading doublestranded (ds) RNA (EA_{dsRNA}), although the physiological relevance of this catalytic activity is unclear. Similar EA_{dsRNA} was found among RNASE 1 of human, rhesus monkey and douc langur, with that of douc langur RNASE 1B reduced to approximately 0.3%. As one interpretation, RNASE 1B can afford to lose EA_{dsRNA} function because the paralogous RNASE 1 retains it. It is likely that some of the adaptive charge-altering substitutions in RNASE 1B are detrimental to EA_{dsRNA} . To determine which of the nine amino acid substitutions in RNASE 1B are responsible for loss of EA_{dsRNA} , site-directed mutagenesis was used to create mutant forms of douc langur RNASE 1, each with one substitution. It was found that eight of the nine substitutions reduce EA_{dsRNA} substantially, whereas the other has a mild and marginally significant effect.

Similarly, it should be investigated if GH I and GH II of common carp have a difference in activity and, if this is the case, which of the six differing amino acids conduce to change of activity.

V.1.2 Strategies for improving the yield of soluble recombinant proteins in E. coli

For the production of the recombinant growth hormones of common carp an *Escherichia coli* expression system could be used. One disadvantage of expressing heterologous proteins in *E. coli*, however, is that proteins are frequently expressed as insoluble, aggregated folding intermediates called inclusion bodies. The refolding of

inclusion bodies is not simple and a very time consuming process during which a lot of recombinant protein could be lost because it is not possible to fold 100% of recombinant protein in inclusion bodies properly. For activity tests of different forms of recombinant proteins it is better to obtain recombinant proteins in soluble native form without the need to refold them to save time and operations.

In most cases, solubility is not an all-or-none phenomenon; the vector, host, and culture conditions combinations can increase or decrease the proportion of soluble and insoluble forms of protein obtained.

Approaches employed to improve the production of recombinant protein in native form include optimization of cultivation conditions. Cultivation parameters are important factors to achieve the native conformations of recombinant proteins. For example, the enzyme α -glucosidase PI from *Saccharomyces cerevisiae* is a soluble, monomeric, cytoplasmic enzyme with a molecular mass of 67,500 Dalton. It is not glycosylated and contains four unlinked cysteine residues. Overexpression of α -glucosidase PI leads, as of most recombinant proteins, to formation of inclusion bodies under optimal growth conditions for *E. coli* at 37°C in complex medium, but was folded properly at the lower temperature of 24°C (Kopetzki, 1989).

It has been proposed that the yield of native protein depends on the rates of folding, aggregation and protein synthesis (Kiefhaber et al., 1991; Rudolph, 1996). Hence, an increase of folding rate and a decrease of protein synthesis rate lead to increased amount of correctly folded protein. Under growth conditions such as low temperature or low gene expression rate by using weaker promoter or partial induction, formation of inclusion bodies can be reduced or avoided (Kopetzki et al., 1989; Schein, 1989; Bowden & Georgiou, 1990; Chalmers et al., 1990; Shatzman, 1990; Strandberg & Enfors, 1991). Low temperature results in higher soluble product concentration and lower amount of inclusion bodies (Schein & Noteborn, 1988; Kopetzki et al., 1989).

Aside from optimization of cultivation parameters the choice of vector and expression host can significantly increase the activity and amount of target protein in the soluble fraction. A vector can enhance solubility and/or folding in one of the next ways: (1) enable fusion to a polypeptide that itself is highly soluble (e.g. glutathione-S-transferase (GST), thioredoxin (Trx), N utilization substance A (NusA)) (Nygren et al., 1994; LaVallie et al., 1993; Zheng et al., 2003); (2) enable fusion to an enzyme that catalyzes disulfide bond formation (e.g. thioredoxin, DsbA, DsbC) (Collins-Racie et al., 1995; Stewart et al., 1998).

When using vectors designed for cytoplasmic expression, folding can be improved in hosts that allow the formation of disulfide bonds in the cytoplasm (e.g. *trx*B and *gor*

mutations). Many proteins require the formation of stable disulfide bonds to fold properly into a native conformation. Without disulfide bonds, these proteins may be degraded or accumulate as inclusion bodies. A limitation of the production of properly folded proteins in *E. coli* has been the relatively high reducing potential of the cytoplasmic compartment; disulfide bonds are usually formed only upon export into the periplasmic space. Bacterial strains with glutathione reductase (*gor*) and/or thioredoxin reductase (*trxB*) mutations enhance the formation of disulfide bonds in the *E. coli* cytoplasm (Prinz et al., 1997; Aslund et al., 1999; Bessette et al., 1999).

V.1.3 Strategies for improving the purification of recombinant proteins

Fusion tags can facilitate detection and purification of the target protein. The His-Tag sequence is very useful as a fusion partner for purification of proteins in general. Immobilized metal-ion affinity chromatography (IMAC) became a useful technique for the separation of proteins when the chelating ligand iminodiacetic acid (IDA) was introduced by Porath et al. (1975). Separation is based on the interaction of histidine, cysteine, and tryptophan side chains with transition metal ions (e.g. Cu^{2+} , Ni^{2+} , Co^{2+} , Zn^{2+}) immobilized via IDA to a porous chromatographic support (Belew & Porath, 1990). The sensitivity of the method for the presence and distribution of these amino acid residues on the surface of the protein explains why relatively few proteins show strong binding to the adsorbents, making IMAC an efficient step in purification protocols (Porath & Belew, 1983).

Smith et al. (1987) found that small peptides containing several histidine and tryptophan residues bind strongly to transition metal ions immobilized on IDA-Sephadex G-25. These authors suggested that insertion of such a chelating peptide at the amino terminus of a recombinant protein could be a general method for the specific capture of the protein by IMAC (Smith & Pidgeon, 1986; Smith et al., 1987). It was subsequently shown that a histidine-tryptophan dipeptide fused to the aminoterminal sequence of recombinant pro-insulin significantly increased the affinity of the protein for Ni^{2+} -IDA-Sephadex G-25 (Smith et al., 1988). Hochuli et al. (1987) showed that an oligopeptide containing two adjacent histidine residues was more strongly adsorbed to Ni^{2+} -NTA (nitrilotriacetate) adsorbent than a peptide with three noncontiguous histidine residues.

Hochuli et al. (1988) showed for the first time that the use of IMAC to separate an expressed recombinant protein fused with a hexa-histidine peptide tag yielded highly purified protein in a single chromatographic step under both native and denaturing conditions. The use of this approach has increased dramatically, and fusion with histidine peptides is now a major technique for tagging recombinant proteins for affinity purification.

V.2 Material and Methods

V.2.1 Experimental fish and collection of samples

The mirror carp fry were obtained from Fischzucht Jänschwalde GmbH fish farm. The fishes were maintained at natural photoperiods and 23-24°C. The fish were fed with commercial dry diet (AQUA VALENT; Märkische Kraftfutter GmbH). The pituitary glands were collected from 1 year old fish and put into TRIzol (Invitrogen). All samples were frozen in liquid nitrogen and kept at -80°C until RNA isolation.

V.2.2 RNA preparation and first strand cDNA synthesis

RNA isolation and first strand cDNA synthesis was carried out as described in Chapter IV sections IV.2.2 and IV.2.3.

V.2.3 Primers and PCR conditions for amplification of GH I and GH II cDNA

Specific primer pairs for cDNA of each gene were constructed, and PCR products with coding sequences of *GH I* or *GH II* were obtained.

Primer pair for *GH I* cDNA:

K-F: CTG AGC GAA ATG GCT AGA GT

K-R: TAC ACC GGT GCC ATC TAC AG

The PCR product should be 656 bp long.

Primer pair for *GH II* cDNA:

CH-F₂: ACC AGG GGA GAG CAT CAG AT

CH-R₂: TGC AGG CAC TGA CTA GCA ATA

The PCR product should be 612 bp long.

Each PCR reaction mix (total 25 µl) was composed of 1 x PCR buffer (20 mM Tris - HCl, (pH 8.8 at 25°C), 10 mM (NH₄)₂SO₄, 10 mM KCl, 0.1% Triton X-100), 4 mM MgCl₂, 0.8 mM dNTPs, 1.0 µM of each primer, 5 µl cDNA and 1.25 units of *Pfu* DNA polymerase (MBI-Fermentas).

The amplification conditions were identical for both primer pairs: 95°C for 3 min then 35 cycles consisting of 95°C for 30 sec, 55°C for 30 sec and 72°C for 1 min.

All PCR products were visualized on 1.7% agarose gels using ethidium bromide staining and UV light. The GeneRuler™ 100 bp DNA Ladder Plus (100 bp - 3000 bp) (Fermentas) served as size standard.

The fragments for *GHI* (656 bp) and for *GHII* (612 bp) were cut from the gel and purified with E.Z.N.A. Gel Extraction Kit (Peqlab Biotechnologie).

V.2.4 Cloning of PCR products from section V.2.3

Purified PCR products were cloned using the pGEM^R-T Easy Vector system (Promega). Primers were constructed for control of inserts in pGEM^R-T Easy-Vector (Promega) by colony PCR and later sequencing:

Vek-1: CGA CTC ACT ATA GGG CGA ATT G

Vek-2: GAC CAT GAT TAC GCC AAG CTA

Each PCR reaction mix (total 25 µl) was composed of 1 x PCR buffer (10 mM Tris - HCl, pH 8.3, 50 mM KCl), 2.5 mM MgCl₂, 0.1 mM dNTPs, 0.2 µM of each primer, 5 µl water with resolved bacteria from one colony and 0.25 units of Taq - DNA polymerase (MBI-Fermentas). The amplification conditions were: 3 min at 94°C, then 35 cycles consisting of 30 sec at 94°C, 30 sec at 55°C, 1 min at 72°C.

After colony PCR some positive clones were selected. The plasmids were isolated from *E. coli* cells using E.Z.N.A. Plasmid Miniprep Kit I (Peqlab Biotechnologie) or Qiagen Plasmid Purification Maxi Kits. After that, plasmids were sequenced to check the insert. The automated sequencing was carried out on Beckman Coulter CEQ 8000 using the CEQ-DTCS Quick Start Kit (Beckman Coulter).

V.2.5 Expression vectors and E. coli strains used for improving the yield of native recombinant protein

Four *E. coli* expression hosts (BL21, BLR, HMS174, Rossetta-gammi) were transformed with expression vectors pET-30 or pET - 32 containing the coding region of *GHI* or *GHII* as inserts. Thereby, eight combinations of expression vectors and hosts were tested for their ability to express in *E. coli* soluble recombinant growth hormone of common carp. Descriptions of used expression hosts are presented in Table V.1 and of used expression vectors in Table V.2.

Table V-1 E. coli expression strains used in present study.

Strains	Genotype and/or description	Source
BL21	F- <i>ompT hsdS_B (r_B- m_B) gal dcm (DE3)</i> General purpose expression host	Novagen
BLR	F- <i>ompT hsdS_B (r_B- m_B-) gal dcm Δ(srl-recA)306::Tn10(Tc^R) (DE3) recA⁻</i> Expression host recommended for use with tandem repeats	Novagen
HMS 174	F- <i>recA hsdR (r_{K12}- m_{K12}⁺) Rif^R (DE3) recA- K12</i> Expression host	Novagen
Rossetta-gammiTM	Δ <i>ara-leu7697 ρlacX74 ΔphoAPvuII phoR araD139 ahpC galE galK prsL F'[lac+(lacIq)pro] gor522::Tn10(TcR) trxB::kan (DE3) pRARE (CmR)</i> General expression host; two mutations in cytoplasmic disulfide reduction pathway enhance disulfide bond formation in E.coli cytoplasm, provides rare codon tRNAs.	Novagen

Table V-2 Tested expression vectors (Novagen).

Vector	Selection	Promoter	Fusion tags	Protease site
pET-30	Kan	T7lac	His.Tag, S.Tag	Thrombin, Enterokinase
pET – 32	Amp	T7lac	Trx.Tag TM , His.Tag, S.Tag	Thrombin, Enterokinase

a Expression vector cleavage

The plasmids were cleaved in a total reaction volume of 50 µL with restriction nuclease Bam HI and Nco I at 37°C during 3-3.5 hours. The amounts of restriction nucleases and buffer were according to the protocol recommended by the manufacturer (Fermentas). The cleavage process was stopped by heating at 80°C for 20 min. The reaction was checked on 1.7% agarose gels using ethidium bromide staining and UV light.

b Dephosphorylation of cleaved vectors

The Calf Intestine Alkaline Phosphatase (CIAP) (Fermentas) was used to catalyze the removal of 5'-phosphate groups from DNA to avoid self-ligation of the vector. About 2 units of enzyme per microgram vector DNA were added to the DNA solution after cleavage and the mixture was incubated at 37°C for one hour. The enzyme was deactivated by incubation at 85°C for 15 min.

V.2.6 Recloning of GH I and GH II coding sequences in expression vectors

a Primers to amplify the GH I and GH II coding sequences from cloning vector and recloning them in expression vectors

A procedure for parallel cloning of genes into expression vectors without restriction digestion of PCR products was used (Shih et al., 2002). This method makes use of sticky-end PCR and directional cloning. This procedure allows producing inserts with sticky-ends of restriction enzymes that will cut the coding sequence itself. In case of the two growth hormone proteins of common carp this method allowed to maintain both coding sequences in the same way to insert them in expression vectors (Figure V.1). Two PCR products from one template should be obtained with two primer pairs.

Listed below are the nucleotide sequences of the four sticky-end PCR primers (A-D) used for recloning the two growth hormone full-length coding sequences with stop codon into *Nco I* (CCATGG) and *Bam HI* (GGATCC) sites of expression vectors pET-30 and pET - 32.

Primers for *GH I* and *GH II* coding sequences:

A: **CATGGG** A TCA GAC AAC CAG CGG CTC TT - 3'

B: G A TCA GAC AAC CAG CGG CTC TT - 3'

C: **GA TCC** TTA CTA CAG GGT GCA GTT GGA ATC - 3'

D: **C** TTA CTA CAG GGT GCA GTT GGA ATC - 3'

The location of restriction site nucleotides in sticky-end PCR primers is marked with italics and bold letters. Primers A and B contain the coding sequence of the first seven amino acid residues in *GHI* and *GHI*. Primers C and D contain the complement of a stop codon sequence (underlined) and the complement of the coding sequence of the last seven amino acid residues of *GHI* and *GHI*.

Primer A and D are used for PCR reactions AD with pGEM^R-T Easy Vector with *GHI* as well as with *GHI* inserts obtained in *V.2.4*. The primers B and C are also used for PCR reaction BC with pGEM^R-T Easy Vector with *GHI* as well as with *GHI* inserts obtained in *V.2.4*. For PCR the proofreading Pfu DNA polymerase (Fermentas) was used. PCR conditions were: 3 min at 94°C, then 35 cycles consisting of 1 min at 94°C, 1 min at 54°C, 2 min at 72°C.

b 5'-end phosphorylation of PCR products

PCR products were purified with E.Z.N.A. Cycle-Pure Kit (Peqlab Biotechnologie). The concentration and purity of PCR products were analyzed at 260 nm and 280 nm on Biophotometer (Eppendorf).

The PCR products AD and BC for the *GHI* coding sequence were adjusted to equal concentration and mixed with each other. The same was done with the PCR products AD and BC for the *GHI* coding sequence. The phosphorylation of PCR products of both mixtures was done. The reaction was carried out in 1 x volume T4 polynucleotide kinase buffer (50 mM Tris - HCl (pH 7.6 at 25°C), 10 mM MgCl₂, 5 mM DTT, 0.1 mM EDTA), 2.5 mM ATP and 20 units T4 polynucleotide kinase (Fermentas) in 50 µl volume. The reaction was incubated for 60-90 min at 37°C.

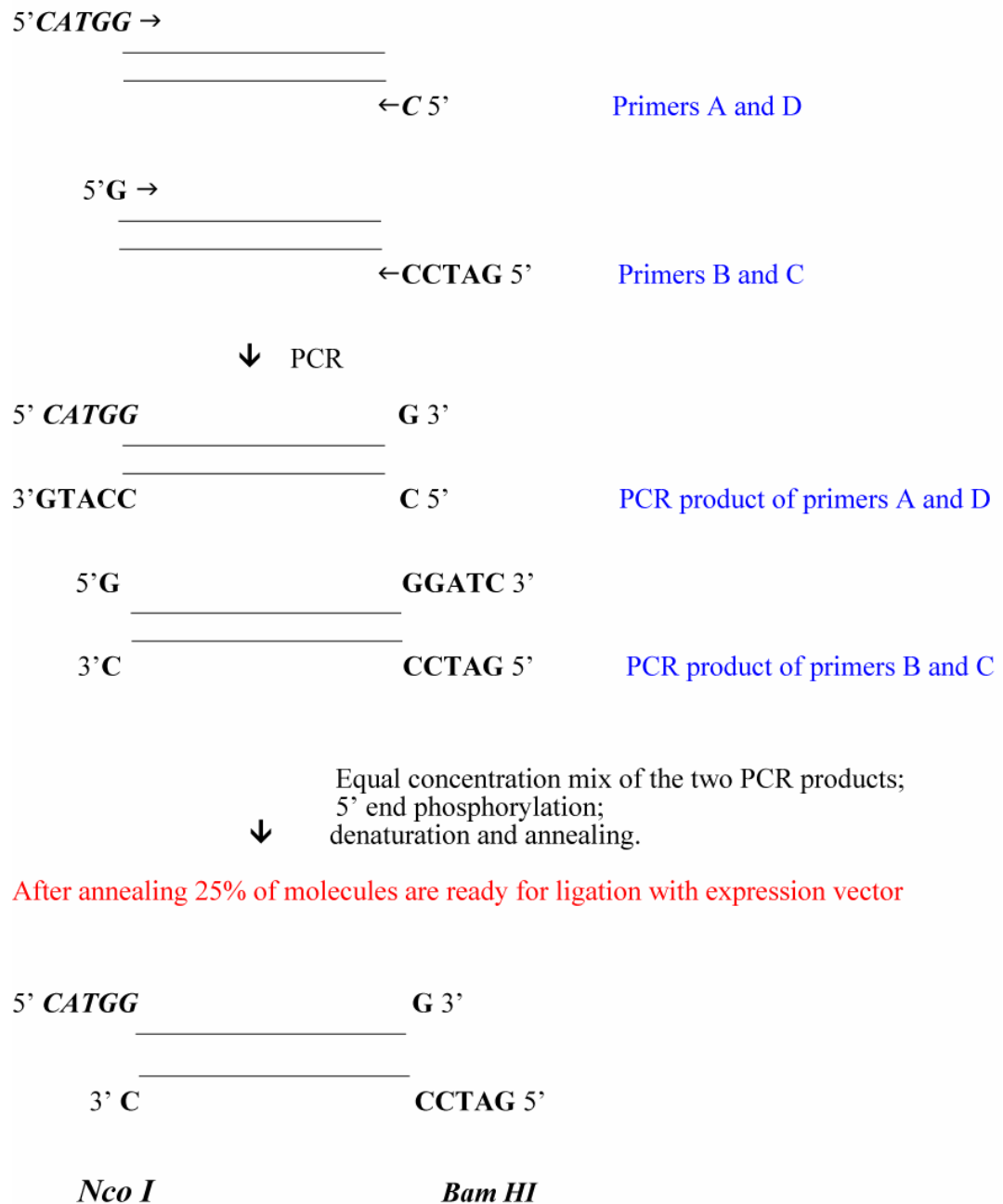


Figure V.1 Molecular cloning strategy: Four PCR primers and two PCRs are used to create double-stranded DNA inserts with cohesive ends that are ready for ligation with expression vectors. *Nco I* - C↓CATGG; *Bam HI* - G↓GATCC restriction sites used to insert *GH I* and *GH II* coding sequences in expression vectors.

c Production of ligation-ready, sticky-end PCR products

To denature the PCR products AD and BC and to inactivate the T4 polynucleotide kinase the reactions were incubated for 5 min at 95°C. Then, the reactions were incubated for 5 min at 65°C to anneal the PCR products. Finally, the reactions were cooled to room temperature (Figure V.1).

d Ligation of expression vectors and sticky-end PCR products

The expression vectors were ligated with sticky-end PCR products in 1 x T4 DNA ligase buffer (40 mM Tris - HCl, 10 mM MgCl₂, 10 mM DTT, 0.5 mM ATP (pH 7.8 at 25°C)) with 5 units T4-Ligase (Fermentas) in 10 µl volume. The amount of fragments was taken in fourfold molar excess relative to the amount of expression vector. The ligation reactions were incubated overnight at 4°C.

V.2.7 Preparation of competent cells

The competent bacteria cells were obtained from Novagen or were prepared as follows: a single colony from an LB plate was used to inoculate 10 ml of LB medium containing the appropriate antibiotic. It was shaken at 225 rpm overnight at 37°C. Then, 100 ml of LB medium containing antibiotic were inoculated with 1 ml of this overnight culture. The cells were grown to OD 0.33 at 37°C.

The cells were harvested by centrifugation at 4500 g for 10 min at 4°C. The supernatant was discarded and cells were resuspended in 100 mM ice cold CaCl₂. After incubation in CaCl₂ bacteria cells could be transformed.

V.2.8 Transformation of the expression vectors with inserts into competent cells

To 20 µl of competent cells 0.2 ng of plasmid was added. The cells were incubated on ice for 5 min. Then, a heat shock was applied at 42°C for 30 sec. The cells were cooled on ice again. After that, 80 µl SOC was added and the reaction mixture was incubated for 60 min at 37°C.

The bacteria were plated on agar plates with appropriate antibiotic.

V.2.9 Cultivation medium

For Luria-Bertani (LB) medium, the following components were dissolved in distilled water:

- bacto trypton 10 g/L
- yeast extract 5 g/L
- NaCl 5 g/L

The Agar plates were composed of:

- bacto trypton 10 g/L
- yeast extract 5 g/L
- NaCl 5 g/L
- Agar 15 g/L

Antibiotics were sterilized by filtration and supplemented aseptically to all media from the agar plates to shake flasks. The final concentrations of the antibiotics were: carbeniciline 50 µg/ml, chloramphenicol 34 µg/ml, tetracycline 25 µg/ml.

V.2.10 Small-scale protein induction

A single colony was picked from a LB plate to inoculate 2 ml of LB medium containing appropriate antibiotic. It was shaken at 225 rpm overnight at 37°C.

5 ml of LB medium containing antibiotic were inoculated with 75 µl of overnight culture. The cells were grown for 3-3.5 hours at 37°C. The culture was cooled to 20°C.

a Preparation of total cell lysates without IPTG induction (negative control 1)

1 ml aliquot from 5 ml culture was transferred into a new culture tube without IPTG. The culture grew for 20-24 hours at 20°C. The cells were harvested by centrifugation at 4500 g for 5 min at 4°C. The supernatant was discarded and cells were resuspended in 100 µl of 2 x SDS sample buffer. The tubes were incubated immediately for 5 min at 95°C.

In all of the following experiments with IPTG induction, the IPTG was added to a final concentration of 1 mM.

b Preparation of total cell lysates after IPTG induction

1 ml aliquot from 5 ml culture was transferred into a new culture tube with IPTG. The culture grew for 20-24 hours at 20°C. The cells were harvested by centrifugation at 4500 g for 5 min at 4°C. The supernatant was discarded and cells were resuspended in 100 µl of 2 x SDS sample buffer. The tubes were incubated immediately for 5 min at 95°C.

c Preparation of soluble protein fraction from total cell lysates after IPTG induction

3 ml aliquot from 5 ml culture was transferred into a new culture tube with IPTG. The culture grew for 20-24 hours at 20°C. The cells were harvested by centrifugation at 4500 g for 5 min at 4°C. The supernatant was discarded and the cell pellet was resuspended completely in 500 µl lysis buffer.

Lysis buffer: (50 mM potassium phosphate pH 7.8, 400 mM NaCl, 100 mM KCl, 0.5% Triton X-100, 10% glycerol, 5 mM Imidazole) with 30 mg/ml lysozyme (Sigma).

The tube was shaken for 2-3 hours at 225 rpm. The lysate was centrifuged at 14000 g for 20 min at 4°C. 100 µl of supernatant was mixed with 100 µl of 2 x SDS sample buffer. The tubes were incubated immediately for 5 min at 95°C.

d Preparation of total cell lysates (cells were not transformed with expression vector) after IPTG induction (negative control 2)

A single colony was picked from a LB plate to inoculate 2 ml of LB containing antibiotic. It was shaken at 225 rpm overnight at 37°C.

5 ml of LB medium containing appropriate antibiotic were inoculated with 75 µl of overnight culture. The cells were grown for 3-3.5 hours at 37°C. The culture was cooled to 20°C. 1 ml aliquot of culture was transferred into a new culture tube with IPTG. The culture grew for 20-24 hours at 20°C.

The cells were harvested by centrifugation at 4500 g for 5 min at 4°C. The supernatant was discarded and cells were resuspended in 100 µl of 2 x SDS sample buffer. The tubes were incubated immediately for 5 min at 95°C.

V.2.11 Large-scale protein induction

2 L LB medium were inoculated with overnight culture from a single colony. The cells were grown to OD of 0.4 at 37°C. The culture was cooled to 20°C and induced with IPTG. Then, the culture was shaken for 20-24 hours at 20°C.

V.2.12 Polyacrylamide gel electrophoresis of proteins

The gel electrophoresis was carried out in a Mini-PROTEAN 3 Cell (Bio-Rad). The resolving gel was Tris/HCl 10% with SDS, prepared according to Laemmli (1970). The 2 x sample loading buffer was composed of 0.150 M Tris/HCl pH 6.8, 20% (v/v) glycerol, 4% (w/v) SDS, 4% (v/v) 2-mercaptoethanol, 0.2% (w/v) bromphenolblue. The total cell fractions or insoluble cell fractions were resuspended in 100 µl loading buffer, soluble

proteins were diluted with one volume loading buffer. After incubation at 95°C for 10 min, the samples were loaded on the gels and were run at 200 V for about 45-50 min in running buffer (25 mM Tris, 192 mM glycine, 0.1% (w/v) SDS). The Protein Molecular Weight Marker (Fermentas) with molecular weights in the range from 14.4 to 116 kDa served as standard.

V.2.13 Polyacrylamide gel staining

Coomassie blue- and silver-staining was used to visualize proteins on polyacrylamide gels. Coomassie blue staining was used for qualitative analysis of proteins. Silver staining was used for visualization of proteins of low concentration and analysis of purity of recombinant proteins.

a Coomassie-blue staining

For staining, gels were rinsed in Coomassie-blue-solution for two hours. Gels were destained using destaining solution, containing 40% (v/v) methanol and 10% (v/v) acetic acid, until the background became colorless.

Staining solution:

- Coomassie Brilliant Blue G 250 (Roth)	1% (w/v)
- Methanol	40% (v/v)
- Acetic acid	10% (v/v)

b Silver staining

Before staining gels were washed thoroughly with deionized water. Then, they were rinsed in aqueous solution containing 0.8% (w/v) silver nitrate (AgNO₃), 0.075% (v/v) NaOH and 0.42% (v/v) ammonium hydroxide for 15 min with gentle, constant agitation. The gels were developed by washing in 100 ml aqueous solution containing 0.5 ml 1% citric acid and 50 µl 38% formaldehyde. The development of the gels was stopped by transferring them to 1% acetic acid solution.

V.2.14 Protein concentration determination

Protein concentrations were estimated with Coomassie Plus-The Better BradfordTM Assay Kit (PIERCE) and procedure was carried out according to the recommendations of the manufacturer. Bovine serum albumin (BSA) (PIERCE) was used as standard. The adsorption at 595 nm was measured on Biophotometer (Eppendorf).

V.2.15 Purification of polyhistidine-containing soluble recombinant protein

The cells from culture described in *V.2.11* were harvested by centrifugation at 10000 g for 10 min. The supernatant was discarded. The precipitated cells were resolved in lysis buffer (see section *V.2.10*). The solution was shaken for 2-3 hours at 225 rpm. The lysate was centrifuged at 15000 g for 20 min to pellet the cellular debris. The supernatant was transferred to a fresh flask. Ammonium sulfate was added to the supernatant to the end concentration of 1.36 M.

The protein precipitate was collected after ammonium sulfate addition by centrifugation at 15000 g for 20 min. The precipitate was redissolved in binding buffer for purification by immobilized metal-ion affinity chromatography (in this case Ni-ion). The HiTrap affinity columns were prepacked with precharged Ni - SepharoseTM (Amersham Biosciences). The purification was carried out on the AKTA FPLC system (Amersham Biosciences) using UNICORN 5.01 software. This system allowed to save chromatograms.

Binding buffer:

20 mM sodium phosphate, 0.5 M NaCl, 20 mM imidazole, pH 7.4.

Elution buffer:

20 mM sodium phosphate, 0.5 M NaCl, 500 mM imidazole, pH 7.4.

The columns were equilibrated with binding buffer. Following loading, the column was washed with binding buffer. After that, gradient wash from 20 mM to 500 mM of imidazole was carried out. The adsorbance at 280 nm was controlled to see the peaks.

a Removing the His-Tag fusion protein

Recombinant Enterokinase Kit (Novagen) was used to remove His-Tag and Trx fusion protein from the recombinant GH proteins. Recombinant Enterokinase (rEK) is a highly purified preparation of the catalytic subunit of bovine enterokinase, which recognizes the identical cleavage site as the native enzyme, AspAspAspAspLys, and has similar enzymatic activity.

The cleavage reaction was carried out according to recommendations of the producer. After cleavage the protein mixture was loaded again on the preequilibrated HiTrap affinity column as described above.

V.2.16 Determination of the monomer content

HPLC gel-filtration chromatography was performed on a Nucleosil 125-5 GFC column (Goehler) in Gilson apparatus. The column was preequilibrated with 0.05 M

NaH₂PO₄ and 0.25 M NaCl at pH 7. The column was developed in the same buffer at a rate of 1.0 ml/min at room temperature. Protein content was monitored by absorbance at 280 nm. The retention time of several known proteins, such as bovine serum albumin, hGH and β -lactoglobuli was determined to calibrate the column.

V.2.17 Recombinant common carp GH activity test

Since IGF - I plays a major role in the endocrinological regulation of growth and because GH acts via that factor, expression of IGF - I was studied in ovaries and liver of the tilapia species *Oreochromis mossambicus* (Schmid et al., 1999a; Schmid et al., 2000; Schmid et al., 2003). A method for preparation of primary liver cell cultures from teleosts was developed by Schmid et al. (1999b). Also, the dependence of IGF - I expression from recombinant tilapia growth hormone was demonstrated in primary liver cell culture of tilapia (Schmid et al., 2000). Based on this method, the activity of recombinant common carp GH could be tested as its ability to stimulate the IGF - I mRNA signal in primary cultured hepatocytes from common carp.

a Isolation of hepatocytes

Liver of common carp of approximately 300-500 g body weight were perfused *in situ* to isolate hepatocytes according to Schmid et al. (2000). The animals were injected with a single dose of 3000 U heparin (Roth) dissolved in 0.5 ml dH₂O, and anaesthetized by immersion in water containing 0.1% MS 222 (Sigma). A catheter (1 mm diameter) was introduced into the bulbus arteriosus and the liver perfused retrogradly with 100 ml calcium-magnesium-free (CMF) medium (Tab. V.3) to wash out the blood. The tissue was digested with 50 ml calcium-magnesium containing (CMC) medium (Tab. V.3) supplemented with collagenase D (0.5 mg/ml) (Roche). The liver was exposed to the collagenase D for 20 min. Finally, the liver was perfused with 50 ml CMF medium, excised, transferred into ice-cold CMF medium and minced. The cells were selected by filtration through nylon gauzes with 250, 100 and 50 μ m meshes. Dispersed cells were washed and collected by two different centrifugation steps (5 min at 700 rpm, 5 min at 500 rpm) using ice-cold CMF medium. The supernatant was discarded each time and the final cell pellet resuspended in minimal essential medium (MEM) (Tab. V.3). Cells were counted and the number of cells was adjusted to approximately 1.5×10^6 /ml MEM. Aliquots (1.5 ml) of cell suspension were seeded as a monolayer into sterile plastic Petri dishes (35 mm diameter) (Primaria, Falcon). The cultures were incubated at 20°C under high humidity with normal air in an incubator (Heraeus).

All media were adjusted to pH 7.5 and 10000 U penicillin and 10 mg streptomycin per 100 ml medium were added after sterile filtration just before use. All chemicals were obtained from Sigma except for amino acids, vitamins and glutamine, which were purchased from Serva.

Table V-3 Media used for perfusion and culture of common carp hepatocytes.

Component	CMF	CMC	MEM
NaCl (mM)	142	142	142
KCl (mM)	5.4	5.4	5.4
CaCl (mM)	0	2.4	2.4
MgSO ₄ 7H ₂ O (mM)	0	0	0.81
Na ₂ HPO ₄ (mM)	0.42	0.42	0.42
KH ₂ PO ₄ (mM)	0.44	0.44	0.44
NaHCO ₃ (mM)	0.43	0	0.43
HEPES (mM)	15	15	20
Amino acids (50x)	0	0	1:50
Vitamins (100x)	0	0	1:50
NaEDTA (mM)	5	0	0
Glutamine (mM)	0.01	0	0

b Treatment of hepatocytes

In order to test the activity of rcGH produced in the present study the *IGF - I* mRNA level in common carp hepatocytes exposed to the recombinant growth hormone was measured. The concentration of rcGH in medium was 10 nM and 25 nM. Hepatocytes were exposed for 10.5 hours.

As controls, *IGF - I* mRNA level was investigated in primary cultured hepatocytes kept in MEM without any further treatment. At 0 and 10.5 hours from control dishes and at 10.5 hours from exposed dishes the medium was removed and the cells were lysed in 500 µl TRIzol (Invitrogen). After that, the cells were frozen in liquid nitrogen. Samples were stored at -80°C until required for further processing.

c Semiquantitative determination of the IGF - I signal using RT-PCR

Total RNA was extracted and cDNA synthesis was carried out as described above in Chapter IV section IV.2.2 and IV.2.3. For analysis of the IGF - I signal dependency in primary cultured hepatocytes from rcGH in medium, the cDNA sequence of common carp IGF - I (GenBank accession number D83271) was taken to construct specific primers. This cDNA sequence is the transcript of one of the two IGF - I genes of common carp (see Chapter III of the present study). There are no data if the duplicated IGF - I genes of common carp have the same or different expression pattern. Therefore, the mRNA level of only one IGF - I gene was determined in the present study to minimize the mistakes that could arise from a possibly different regulation of the duplicated common carp IGF - I genes. The primers were specific for the transcript of only one IGF - I gene. The β -actin signal served as an internal standard in the experiment. Only one sequence for β -actin (M24113) of common carp was available in GenBank. The primer pairs for IGF - I (Fas + Ras) and β -actin (F β -ac + R β -ac) were constructed using Primer 3 software (Rozen & Skaletsky, 2000):

Fas: ACA GTC CCA GGA CAC CAA AG

Ras: CAA GGG TTC CAA ACG GTC TA

F β -ac: AAG GCC AAC AGG GAA AAG AT

R β -ac: TAC CGC AAG ACT CCA TAC CC

PCR reaction mixes (25 μ l) consisted of: 1 x PCR buffer (10 mM Tris - HCl, pH 8.3, 50 mM KCl) with 2.5 mM MgCl₂, 5 μ l cDNA, 0.2 μ M of each primer, 0.1 mM dNTPs, and 0.25 units of Taq - DNA polymerase (MBI-Fermentas).

The amplification conditions were as follows: 3 min at 95°C, then 30 cycles consisting of 30 sec at 95°C, 30 sec at 54°C, 1 min at 72°C.

PCR fragments were separated on 1.7% agarose gels and stained with ethidium bromide. The optical density of *IGF - I* bands was normalized visually with that of β -actin.

V.3 Results and Discussion

The scheme of production and purification steps for recombinant cGH is shown on Figure A.8 in Appendix. Only the combination of Rossetta-gammi expression host and pET - 32 expression vector with inserts of *GH I* or *GH II* coding region allowed to produce recombinant proteins in soluble fraction in great amount. All other combinations of expression hosts and vectors produced recombinant proteins in inclusion bodies. For all combinations of expression hosts and vectors the protein production was carried out at

20°C. Figure V.2 shows a photo of SDS-polyacrylamide gel after electrophoresis under denaturing conditions with Coomassie-blue staining. On the gel the recombinant protein could be seen in inclusion bodies fraction and in soluble fraction after induction of protein expression with IPTG in Rossetta-gammi expression host with transformed pET - 32 expression vector. The recombinant protein has high expression level and abundance in soluble fraction in comparison to the other proteins (Fig. V.2, lane 3). In addition, two negative controls were checked on the gel: (1) protein extract from Rosetta-gammi expression host with expression vector pET - 32 but without IPTG induction (Fig. V.2, lane 6) and (2) protein extract from Rosetta-gammi expression host without expression vector pET - 32 but with IPTG induction (Fig. V.2, lane 7).

The common carp growth hormone protein is about 22 kD. Due to the fusion of growth hormone with thioredoxin (Trx) and HisTag in the pET - 32 expression vector the theoretical molecular weight of the resulting target protein is about 39 kD. In lanes 2, 3 and 4 of Figure V.2 a very intensive band of about this molecular weight could be seen. In control lanes 6 and 7 this band was not found with such great intensity. This means that recombinant growth hormone was expressed in Rosetta-gammi transformed with pET - 32 expression vector containing the coding sequence of common carp GH and the expression induced by IPTG (Fig. V.2, lanes 2 and 3).

The successful production of recombinant common carp growth hormone in great amount in soluble fraction in combination of Rosetta-gammi expression host and pET - 32 expression vector could be explained by attributes of these expression vector and expression host. The pET - 32 expression vector enables the fusion of thioredoxin to growth hormone. Thioredoxin is a highly soluble protein and in addition is an enzyme that catalyzes disulfide bond formation. On the other hand, only Rosetta-gammi (unlike other bacterial hosts used in the present studies) has two mutations in the cytoplasmic disulfide reduction pathway. There are glutathione reductase (*gor*) and thioredoxin reductase (*trxB*) mutations enhancing disulfide bond formation in *E. coli* cytoplasm. Therefore, due to the combination of favourable attributes of expression host and vector recombinant protein could be obtained in great amount in soluble fraction.

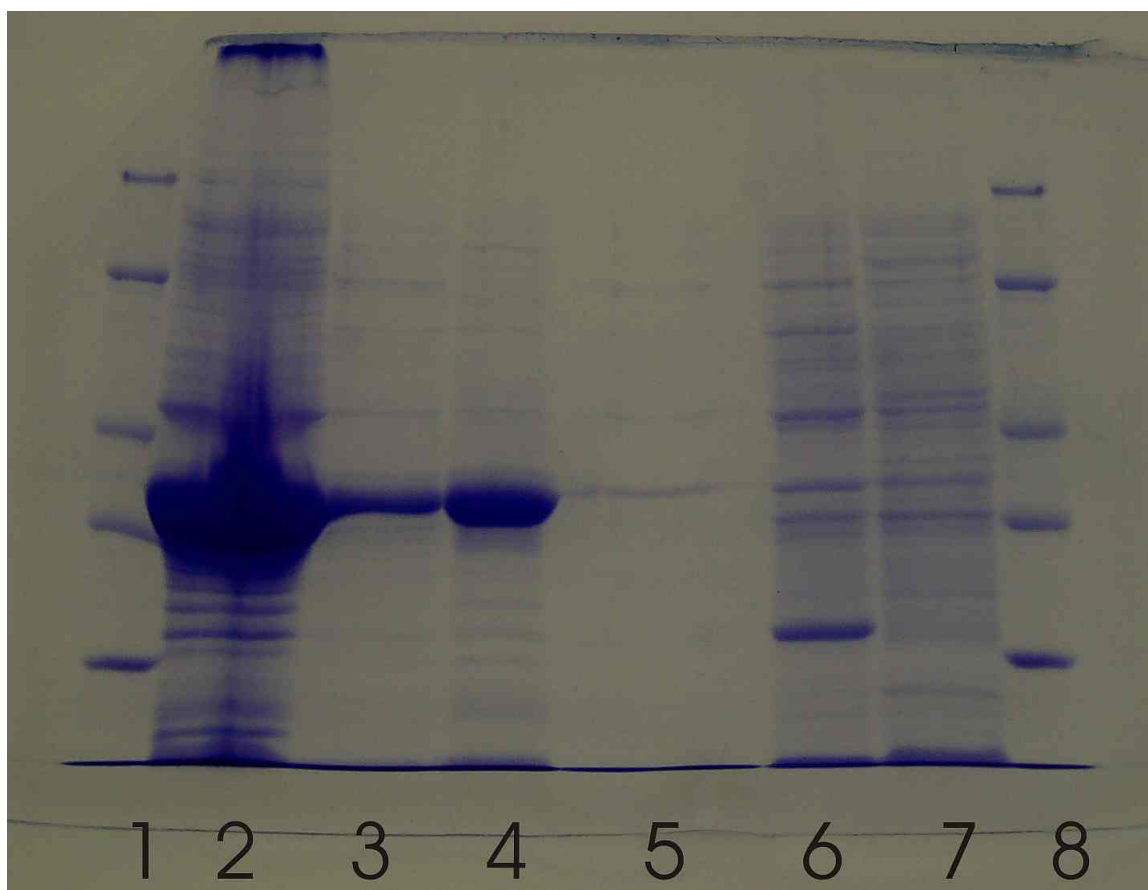


Figure V.2 Expression of recombinant protein in Rossetta-gammi cells transformed with pET - 32. Induction was performed by IPTG. Gel was stained with Coomassie brilliant blue. Lanes 1 and 8, protein molecular weight markers (from top to bottom in kDa): 116, 66, 45, 35, 25; Lane 2, nonsoluble pellet after cell lysis; Lane 3, soluble protein fraction in the supernatant after cell lysis; Lane 4, precipitate of proteins after concentration from supernatant (Lane 3) with ammonium sulfate; Lane 5, proteins remained in supernatant after concentration from supernatant (Lane 3) with ammonium sulfate; Lane 6, Rossetta-gammi transformed with expression vector pET - 32 without induction by IPTG; Lane, 7 Rossetta-gammi without expression vector induced by IPTG.

The proteins from soluble fraction were concentrated with ammonium sulfate. The precipitate was dissolved in 20 mM sodium phosphate buffer and put on the HiTrap affinity column with Ni - Sepharose in the AKTAFPLC system (Amersham Biosciences). Through gradient washes with imidazole from 20 mM to 500 mM in the buffer untagged, contaminating proteins were removed. The purity of the fusion protein that is composed of HisTag, thioredoxin and growth hormone was significantly increased (Fig. V.3, lanes 2 and 4).

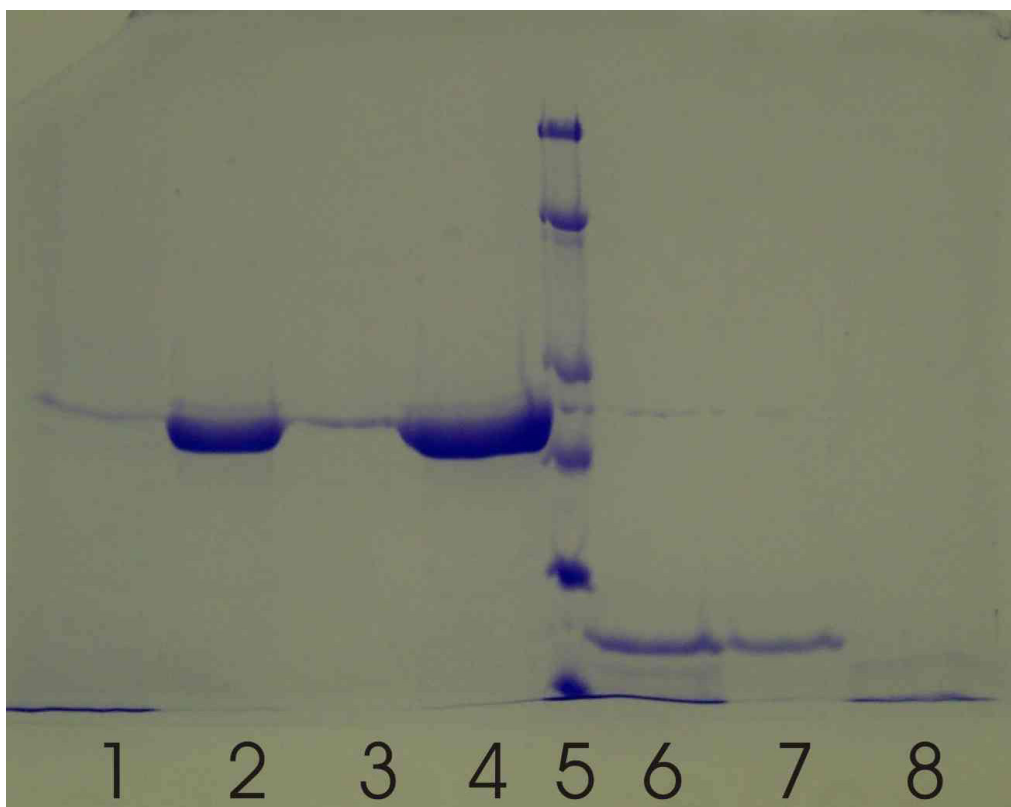


Figure V.3 Electrophoretic analysis of recombinant protein on a 10% SDS-PAGE under denaturing reducing conditions at different stages of purification. Gel was stained with Coomassie brilliant blue. Lanes 1 and 3, first separation on the HiTrap affinity column with Ni - Sepharose, the first peak containing untagged, contaminating proteins (Fig. V.4); Lanes 2 and 4, first separation on the HiTrap affinity column with Ni - Sepharose, the second peak containing fusion of growth hormone protein with highly soluble protein thioredoxin (Trx) and HisTag (Fig. V.4); Lane 5, protein molecular weight marker (from top to bottom in kDa):116, 66, 45, 35, 25, 18; Lane 6, protein mixture after cleavage of the fusion protein with Enterokinase; Lane 7, second separation of protein mixture on the HiTrap affinity column with Ni - Sepharose after cleavage with Enterokinase, the first peak containing recombinant growth hormone (Fig. V.5); Lane 8, second separation of protein mixture on the HiTrap affinity column with Ni - Sepharose after cleavage with Enterokinase, the second peak containing fusion of thioredoxin and HisTag and contaminating proteins with high affinity to Ni - Sepharose (Fig. V.5).

During the first separation on the HiTrap affinity column with Ni - Sepharose the fusion protein was in the second peak that appeared from 60 mM to 260 mM imidazole in buffer (Fig. V.4). The affinity of fusion protein to Ni - Sepharose is high through six adjacent histidine residues (HisTag) (Smith & Pidgeon, 1986; Smith et al., 1987; Hochuli et al., 1987; Hochuli et al., 1988). Therefore, higher concentrations of imidazole should be used to remove the fusion protein from the column than for other proteins. This property

helps to separate an expressed recombinant protein with high purity already in a first chromatographic step.

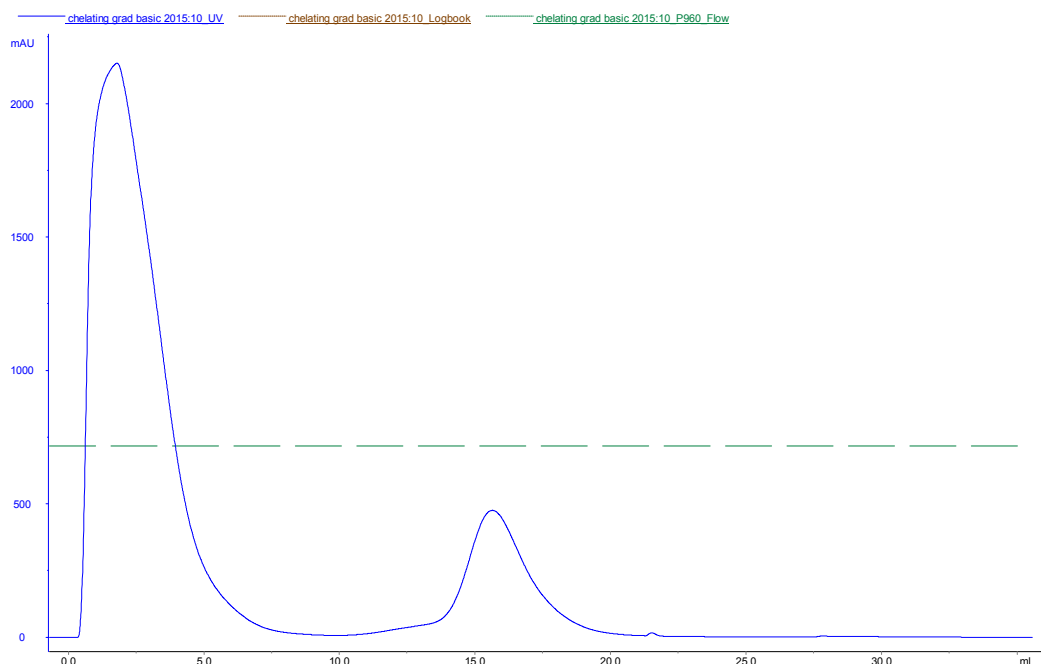


Figure V.4 The first separation on the HiTrap affinity column with Ni - Sepharose. The column was equilibrated with 20 mM sodium phosphate buffer with 20 mM imidazole, pH 7.4. The precipitate of proteins (after ammonium sulfate concentration) containing fusion of growth hormone with high soluble protein thioredoxin (Trx) and HisTag was solubilized in 20 mM sodium phosphate buffer with 20 mM imidazole and applied to the column. The column was washed with the same buffer and untagged, contaminating proteins were removed - the first peak (20 mM imidazole). Then, elution was carried out in the same buffer using a gradient with imidazole from 20 mM to 500 mM. The second peak was eluted with 60-260 mM of imidazole. It contained fusion of growth hormone with high soluble protein thioredoxin (Trx) and HisTag.

After the first separation on the HiTrap affinity column the cleavage of fusion protein to remove the thioredoxin with HisTag from growth hormone was carried out. After cleavage with Enterokinase, Enterokinase was captured with EKapture Agarose (Recombinant Enterokinase Kit, Novagen). Then, the protein mixture was separated on HiTrap affinity column again with the same buffer and wash conditions as in first separation. The recombinant growth hormone was eluted in the first peak with 20 mM imidazole concentration (Fig. V.5). The fusion of thioredoxin with HisTag was eluted in the second peak with imidazole concentration from 60 to 260 mM in buffer. Without fusion to HisTag the recombinant growth hormone has a very weak affinity to Ni - Sepharose and consequently could easily be separated from fusion protein of thioredoxin and HisTag and remaining contaminating proteins with high affinity to Ni - Sepharose.

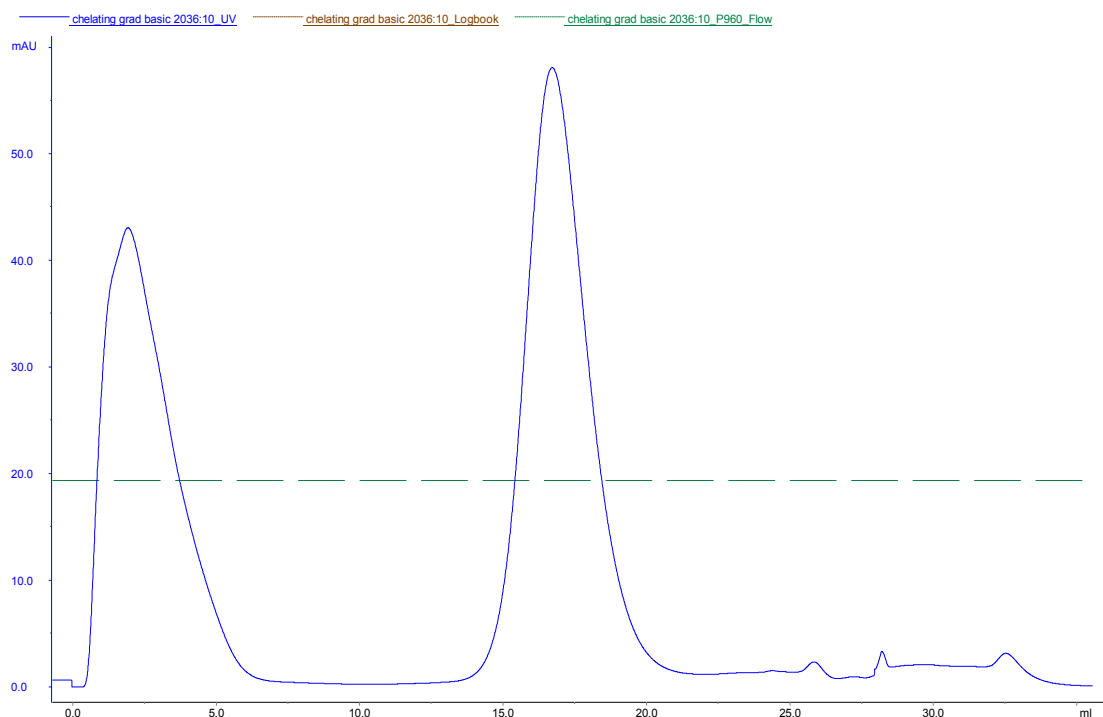


Figure V.5 The second separation on the HiTrap affinity column with Ni - Sepharose. The separation was done after cleavage with Enterokinase of fusion protein consisting of growth hormone and thioredoxin with HisTag. The column was equilibrated with 20 mM sodium phosphate buffer containing 20 mM imidazole, pH 7.4. The mixture of proteins containing the recombinant growth hormone, thioredoxin (Trx) and HisTag connected with each other and other contaminating proteins with high affinity to Ni - Sepharose was applied to the column. The column was washed with the same buffer and recombinant growth hormone was eluted in the first peak (20 mM imidazole). Then, elution was carried out in the same buffer using a gradient with imidazole from 20 mM to 500 mM. The second peak was eluted with 60-260 mM of imidazole. It contained thioredoxin (Trx) connected with HisTag and contaminating proteins with high affinity to Ni - Sepharose.

The purity of rGH was very high after all purification steps (Fig. V.6, lane 7). HPLC of recombinant growth hormone of common carp on Nucleosil 125-5 GFC column showed that it consisted of 95% monomeric GH, having retention time equal to that of recombinant hGH.

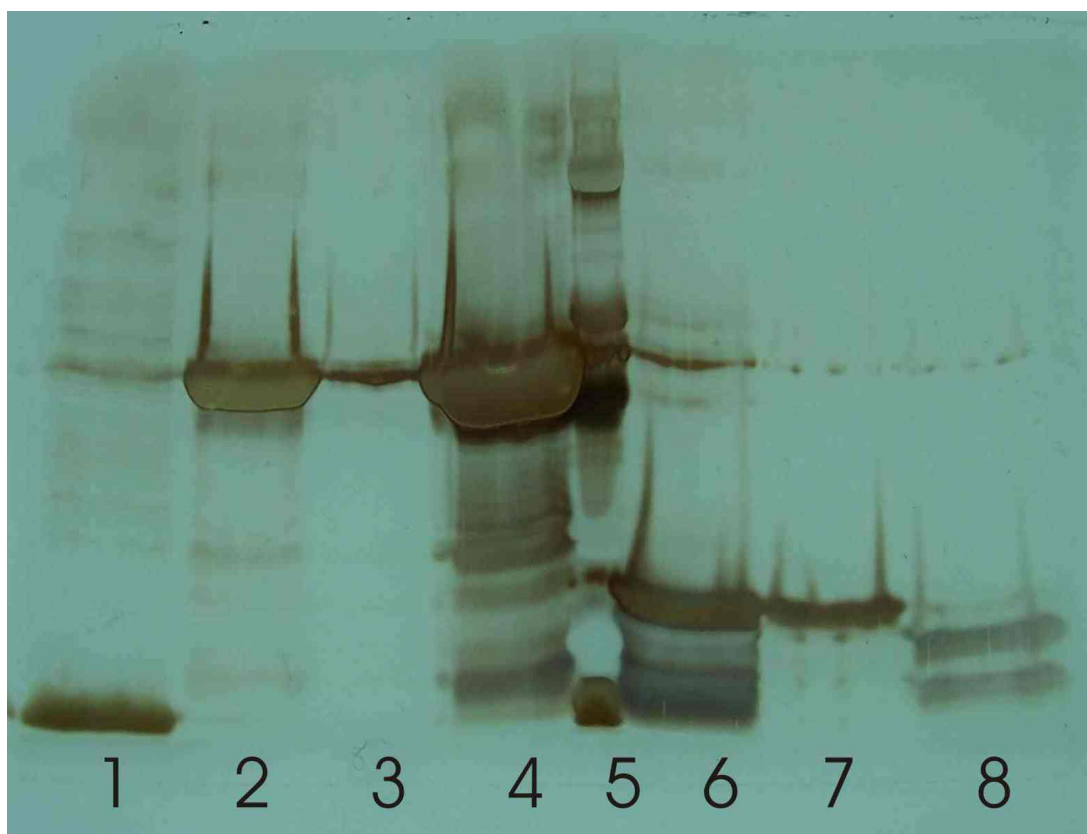


Figure V.6 Electrophoretic analysis of recombinant protein on a 10% SDS-PAGE under denaturing reducing conditions at different stages of purification. Gel was stained with silver, the staining is very sensitive. Although the sensitivity varies somewhat among different proteins, generally 2 ng could be readily detected. Lane 5, protein molecular weight marker (from top to bottom in kDa): 116, 66, 45, 35, 25, 18; Lane 4, precipitate of proteins after concentration with ammonium sulfate from supernatant after cell lysis; Lane 3, proteins remained in supernatant after concentration with ammonium sulfate from supernatant after cell lysis; Lane 1, first separation on the HiTrap affinity column with Ni - Sepharose, the first peak containing untagged, contaminating proteins; Lane 2, first separation on the HiTrap affinity column with Ni - Sepharose, the second peak containing fusion of growth hormone protein with high soluble protein thioredoxin (Trx) and HisTag; Lane 6, protein mixture after cleavage of the fusion protein (fusion of growth hormone with thioredoxin (Trx) and HisTag) with Enterokinase; Lane 7, second separation of protein mixture on the HiTrap affinity column with Ni - Sepharose after cleavage with Enterokinase, the first peak containing recombinant growth hormone; Lane 8, second separation of protein mixture on the HiTrap affinity column with Ni - Sepharose after cleavage with Enterokinase, the second peak containing fusion of thioredoxin with HisTag and remaining proteins with high affinity to Ni - Sepharose.

In summary, the two step purification of recombinant growth hormone on HiTrap affinity column with Ni - Sepharose in first step as fusion with HisTag and Trx and in second step without HisTag allows to produce recombinant protein of high purity. In the

first step all contaminating proteins with low affinity to Ni - Sepharose were removed and in the second step all contaminating proteins with high affinity to Ni - Sepharose.

A qualitative activity test of one of the two produced common carp recombinant growth hormones was done. The activity of rcGH II was tested on common carp primary cultured hepatocytes. The hepatocytes were incubated in medium containing the common carp recombinant growth hormone at doses of 10 nM and 25 nM for 10.5 hours (Fig. V.7). The signal of IGF - I at 25 nM of rGH in medium is stronger than in control samples and than in sample with 10 nM of rGH in medium.

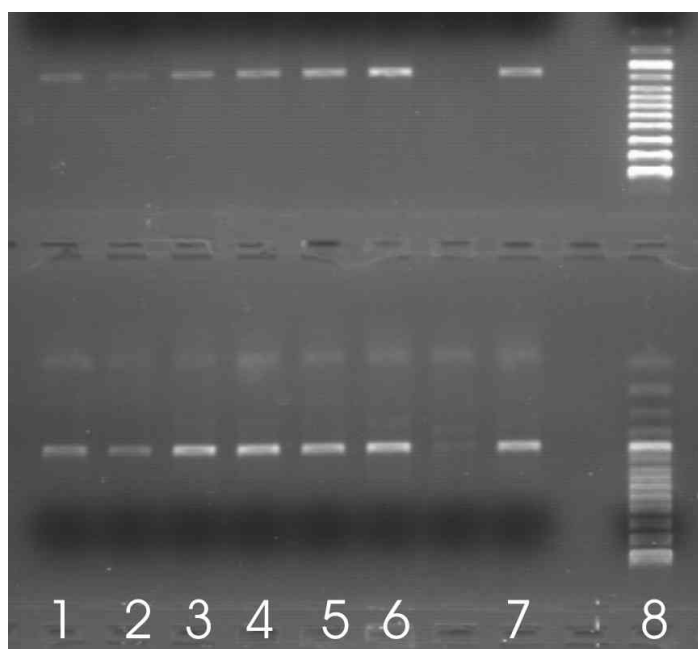


Figure V.7 Analysis of dependency of the IGF - I signal (top part of gel) on the rcGH concentration in the culture medium of primary hepatocytes using semiquantitative RT-PCR. The β -actin signal served as an internal standard in experiment (bottom part of gel).

Lanes 1 and 2, control at 0 hours without rcGH II

Lanes 3 and 4, control at 10.5 hours without rcGH II

Lane 5, hepatocytes exposed to 10 nM of rcGH II for 10.5 hours

Lane 6, hepatocytes exposed to 25 nM of rcGH II for 10.5 hours

Lane 7, hepatocytes exposed to 25 nM of lyophilised rcGH II for 10.5 hours

Lane 8, Molecular weight marker Gene Ruler 100 bp DNA Ladder Plus (MBI-Fermentas)

The simple protein production method developed in the present study could be used to produce active recombinant common carp growth hormone proteins in large amount in few steps and in very short time. This method allows obtaining the recombinant protein directly in soluble form without following necessity to refold it. The refolding of

recombinant proteins is usually a time consuming process. Furthermore, during refolding large amount of recombinant protein could be lost that results in lower yield. The developed method could be employed for analysis if the two growth hormones of common carp have the same or a different activity *in vitro* and *in vivo*. If differences in activity should be found the method could also be used for fast production of a lot of intermediate forms of the two common carp growth hormones to analyse which and how many of the six different amino acids could change the activity.

In addition, it could be tested in future if the two expressed common carp *IGF - I* genes that also demonstrated different evolution rates (see Chapter III of the present study) show the same signal intensity in dependence on common carp rcGHs in hepatocyte culture medium.

Chapter VI Conclusions

In the present study on the duplicated *GH* genes of the tetraploid common carp differences between all introns of both genes were detected. The first introns of the genes are different from each other in nucleotide sequences, predicted DNA curvature and secondary structure. The studies of Bornstein et al. (1988), Franklin et al. (1991), Sica et al. (1992) and Chrysogelos (1993) demonstrate that the first introns of genes can have an important regulatory function. Therefore, it should be tested in future if the observed differences in first introns between the two *GH* genes of common carp could conduce to differences in regulation of both genes.

As it was discovered in the present study, the two *GH* genes of common carp have been subfunctionalized. During the growth of fry mRNA levels of both genes are equal to each other. In contrast, in one year old fish and in three years old males or females the mRNA levels of *GHI* were higher than of *GHI*. This subfunctionalization could be associated with sexual maturation and gonad development. In addition, the mRNA level of *GHI* was higher than of *GHI* in fish adapted to cold temperature and starvation. This different gene regulation could be associated with catabolic state and regulation of energy homeostasis. During the catabolic state IGF - I levels in plasma decrease in concurrence with a reduced expression of hepatic *IGF - I* and *GH receptor* genes (Renaville et al., 2002; Fox et al., 2006).

The division of gene expression after duplication is one form of subfunctionalization. Force et al. (1999) and Lynch & Force (2000) found that, if the original gene had several functional or regulatory domains, deleterious mutations disrupting different subfunctions could accumulate in each of the gene copies. As mentioned above the first introns of the duplicated *GH* genes of common carp differ from each other in sequence, secondary structure and DNA curvature. Secondary structures in introns could play an important role for splicing or affect the accessibility of selected sequences by basic splicing factors (Kirby et al., 1995; Buratti & Baralle, 2004). The sequence-dependent DNA curvature is known to play an important role in the initiation of transcription of many genes (Trifonov, 1985; Hagerman, 1990; Harrington, 1992; Perez-Martin et al., 1994). It would be interesting to analyse which role not only the changes of sequences but also the changes of presence or size of secondary structures and sequence-dependent DNA curvature in the first introns could play for the subfunctionalization of the duplicated *GH* genes of common carp. It could be possible that changes of DNA conformation in non-coding regions of duplicated genes lead to sharing of the duplicated genes expression pattern.

Sequences of promoter regions also play an important role in gene regulation. The promoters of both *GH* genes of common carp were not sequenced in the present study. This should be done for future analyses how differences in promoter regions of both genes could affect the gene sharing pattern of duplicated *GH* genes in common carp.

The equal level of *GH I* and *GH II* mRNA in juvenile fish detected in the present study demonstrates that both genes are important for growth and development of common carp fry. However, it should be noted that in the present study not only pituitary glands were used to determine the *GH II/GH I* mRNA level ratios but whole fry (all somatic tissue) for stages 15 and 22 days after hatching, whole heads for stages 47 and 63 days and whole brain for stage 85 days. It was demonstrated in fish, birds and mammals that the gene encoding for growth hormone is expressed not only in the pituitary gland but also in many other tissues (brain, lymphocytes, placenta, mammary tissue, pineal gland, spermatogonia, primary spermatocytes, oocytes and granulose cells), and GH appears to regulate a host of cellular events through paracrine or autocrine interactions (Izadyar et al., 1999; Butler & Le Roith, 2001; Harvey et al., 2004; Yada et al., 2005). Several investigators reported that the presence of GH or mRNA encoding for this hormone in fish embryos prior to the emergence of a functional pituitary gland suggests that the hormone is not produced by elements of only the somatotrophic axis (Funkenstein & Cohen, 1996; Yang et al., 1999; de Jesus et al., 2002). Therefore, for a better understanding of the functional role of the two *GH* genes in common carp the *GH II/GH I* mRNA level ratios should also be measured separately in each other tissue that could potentially express the growth hormone. In addition, the *GH II/GH I* mRNA level ratios should be examined in common carp embryonic tissues.

The large insertion/deletion of 341 bp in the third intron of the *GH I* gene detected in the present study could impact the intrinsic DNA curvature of this intron. The examination on the extent of this polymorphism has shown that carp populations from Europe and East Asia differ significantly in the frequencies of the short allele. All five domesticated populations from Europe were fixed for the long allele. The domesticated population from China showed the highest frequency of the short allele and was the only one population in which a homozygote for the short allele could be detected. It was demonstrated by Moav et al. (1975) and Wohlfarth et al. (1975) that in poor conditions the fastest relative growth was shown by Chinese carp and in improved environments European carp showed the better growth. Since preliminary analysis with computer software demonstrated that the 341 bp deletion in the third intron leads to a less curved DNA than without deletion this change of DNA conformation could possibly result in different growth of Chinese and European common carp under poor and improved environmental conditions. Even a small intrinsic

curvature may greatly enhance the affinity of protein-DNA contacts (Suzuki & Yagi, 1995). DNA curvature or looping could bring together components of the transcriptional complex that are distant along the DNA sequence (Matthews, 1992). In addition, some authors have demonstrated that polymorphisms in introns of some genes could affect gene regulation (Rosatto, et al., 1999; Van Laere et al., 2003; Johnson et al., 2005). Thus, in future an analysis should also be done to test if the large insertion/deletion in the third intron of the *GHI* gene could affect the regulation of this gene.

Phylogenetic analyses were done with duplicated *GH* genes as well as with other common carp and goldfish loci. Phylogenetic tree topologies for *GH*, *c - myc*, *POMC*, *GH receptor*, prolactin receptor and melanocortin 5 receptor genes demonstrated that these gene duplications occurred already in the ancestor of common carp and goldfish. The common carp and goldfish genes belong to one monophyletic group and only one of two common carp duplicated genes is orthologous to one of two or a single sequence from goldfish. These phylogenetic tree topologies correspond with a tetraploid creation of the common carp and goldfish ancestor. Unfortunately, there are no completed gene linkage maps for common carp and goldfish. Therefore, it is not possible to proof if the studied duplicated genes are unlinked.

Hughes & Hughes (1993) found that amino acid changes were rare relative to the rate of silent substitutions in both copies of 17 duplicated pairs of genes in the anciently tetraploid (30 Myr) frog *Xenopus laevis*. This suggests that neither copy of the gene has been free to accumulate amino acid altering mutations. Relative rate tests demonstrated that both copies of the duplicated gene pairs evolve with the same evolution rate in *Xenopus laevis*. In the present study amino acid changes of duplicated loci in common carp and goldfish were also rare relative to the rate of silent substitutions. On the other hand, some paralogues with increased evolution rates could be identified by the relative rate test of Tajima (1993). The duplicated genes of *Xenopus laevis* were compared with orthologues of human in the relative rate test. In contrast, in the present study the common carp and goldfish paralogues were compared with zebrafish orthologues. Using the human orthologues as outgroup sequences in relative rate tests for fish or amphibian paralogues could possibly conduce to overlooking the differences in evolution rates between paralogues because of the large evolutionary distance between amphibians and mammals or fish and mammals.

Many of the duplicated gene pairs examined in the present study did not show an increased evolution rate in one of the two copies. However, by evaluating whether non-synonymous mutations occur in such a way as to change protein hydrophobicity, charge or

polarity to a greater extent than is expected under random substitution (Hughes et al., 1990) relaxed functional constraints and even evidence of positive Darwinian selection were detected in one of the two copies. Zhang (2003) stated that the current statistical methods for the detection of selection on the sequence analysis level are not efficient enough yet. Therefore, for the analysis of duplicated gene evolution on the DNA level several different methods to sift the natural selection should possibly be used at the same time.

Hughes (1999) assumes that if the ancestral gene already had dual functions gene duplication provides the opportunity for each daughter gene to adopt one ancestral function and further substitutions under positive selection can refine these functions. In addition, many authors show that the number of amino acid substitutions that are involved in a change of gene functions is usually small, so that statistical tests of positive selection are not always powerful (Perutz, 1983; Asenjo et al., 1994; Newcomb et al., 1997). The analysis of He & Zhang (2005) revealed rapid subfunctionalization of duplicated genes in the yeast and human genomes. They expect that subfunctionalization is accompanied by prolonged and substantial neofunctionalization in large proportions of duplicate genes. Partial functional relaxation caused by a loss of ancestral functions provides the opportunity for advantageous mutations which can lead to new functions.

In polyploid species a high level of duplicate gene preservation occurs (Ferris & Whitt, 1979; Nadeau & Sankoff, 1997; Amores et al., 1998; Wendel, 2000). It may be reconciled if dosage requirements play an important role in the selective environment of gene duplicates. Polyploidization preserves the necessary stoichiometric relationships between gene products which may be subsequently maintained by stabilizing selection. On the contrary, after a singleton gene duplication duplicates are out of balance with their interaction partners and may be actively opposed by purifying selection. The level of gene preservation after singleton gene duplications is low. The present data demonstrate that one copy in duplicated gene pairs of common carp and goldfish could have an increased evolution rate, relaxation of functional constraints and even show evidence of positive Darwinian selection. Possibly, tetraploidization that conduces to a longer preservation of duplicated genes in the genome provides more chances for paralogues to subfunctionalize and later to have the opportunity for advantageous mutations which can lead to new functions.

Ohno (1970) has hypothesized that the increased complexity and genome size of vertebrates has resulted from two rounds of whole genome duplications occurring in early vertebrate evolution that provided the requisite raw materials. Conflicting analyses have made this view very controversial now (Lundin, 1993; Spring, 1997; Meyer & Schartl,

1999; Lahrhammar et al., 2002; Guigo et al., 1996; McLysaght et al., 2002; Friedman & Hughes, 2001, 2003). Common carp and goldfish have the same tetraploid ancestor and duplicated genes in these fish species could evolve in different ways as the present study shows. A more precise study how duplicated genes of common carp and goldfish are evolving could probably help to answer the question how whole genome duplication could conduce to the formation of new species and if genome duplication could lead to complexity in vertebrates. The evolution on the DNA level of other paralogues of common carp and goldfish should be analysed. It should be compared how different the evolution of the same gene paralogues is in these two fish species to answer the question how this process could support species differentiation and specialization. More detailed molecular and functional studies should be carried out. The expression patterns of duplicated genes and the biological activity of their protein products should be tested. In addition, complete gene linkage maps should be created for common carp and goldfish.

The concept of the online PCR assay developed in the present study could also be useful for the fast analysis of expression sharing between other duplicated genes (especially in tetraploid species). This approach allows analysing the expression rates of two duplicated genes without using a housekeeping gene that could be very difficult to find in species of tetraploid origin.

The simple protein production method also developed in the present study could be used to produce active recombinant common carp growth hormone proteins in large amounts in a few steps and in very short time. This method allows obtaining the recombinant protein directly in soluble form without following necessity to refold it. The refolding of recombinant proteins is usually a time consuming process. Furthermore, during refolding large amounts of recombinant proteins could be lost resulting in lower yields. The developed method could be employed for analysis if the two growth hormones of common carp have the same or a different activity *in vitro* and *in vivo*. If differences in activity should be found the method could be used for the fast production of a lot of intermediate forms of the two common carp growth hormones to analyse which and how many of the six differing amino acids could change their activity.

In addition, it could be tested if the two expressed common carp *IGF - I* genes that also demonstrated different evolution rates show the same response in dependence on the exposure to common carp recombinant growth hormones in hepatocyte culture media.

All aspects discussed so far are related to basic research on the structure, evolution, expression and function of the duplicated common carp *GH* genes. An applied aspect is the possible development of new approaches to the genetic improvement of common carp

growth performance by selective breeding. The *GH* gene polymorphisms detected in the present study could be used to investigate correlations between them and the growth rate of common carp. If such correlations should be verified *GH* genotypes could be used to develop and establish marker assisted selection (MAS) programs. High effective MAS could then replace the still traditional for common carp but generally only low effective mass selection (Moav & Wohlfarth, 1976; Wohlfarth, 1986). This way, classical methods of selective animal breeding could become again an interesting and competitive alternative to the production of fast growing transgenic fish that possess considerable ecological risks and will probably receive only low acceptance by consumers.

Bibliography

- Agellon & Chen, 1986. Agellon, L.B., Chen, T.T., [1986](#). Rainbow trout growth hormone: molecular cloning of cDNA and expression in *Escherichia coli*. *DNA* **5**, [463-471](#).
- Agellon et al., 1988. Agellon, L.B., Davies, S.L., Chen, T.T., Powers, D.A., [1988](#). Structure of a fish (rainbow trout) growth hormone gene and its evolutionary implications. *PNAS* **85**, [5136-5140](#).
- Ahn & Tanksley, 1993. Ahn, S., Tanksley, S.D., [1993](#). Comparative linkage maps of the rice and maize genomes. *PNAS* **90**, [7980-7984](#).
- Amati et al., 1992. Amati, B., Dalton, S., Brooks, M.W., Littlewood, T.D., Evan, G.I., Land, H., [1992](#). Transcriptional activation by the human c-Myc oncoprotein in yeast requires interaction with Max. *Nature* **359**, [423-429](#).
- Amores et al., 1998. Amores, A., Force, A., Yan, Y.L., Joly, L., Amemiya, C., Fritz, A., Ho, R.K., Langeland, J., Prince, V., Wang, Y.L., Westerfield, M., Ekker, M., Postlethwait, J.H., [1998](#). Zebrafish *hox* clusters and vertebrate genome evolution. *Science* **282**, [1711-1714](#).
- Arabidopsis Genome Initiative, 2000. Arabidopsis Genome Initiative, [2000](#). Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* **408**, [796-815](#).
- Argenton et al., 1993. Argenton, F., Vianello, S., Bernardini, S., Jacquemin, J., Martial, A., Belayew, L., Colombo Bortolussi, M., [1993](#). The transcriptional regulation of the growth hormone gene is conserved in vertebrate evolution. *Biochem. Biophys. Res. Commun.* **192**, [1360-1366](#).
- Arkhipchuk 1999. Arkhipchuk, V.V., [1999](#) Chromosome database. [Database of Dr. Victor Arkhipchuk](#).
- Asenjo et al., 1994. Asenjo, A.B., Rim, J., Oprian, D.D., [1994](#). Molecular determinants of human red/green color discrimination. *Neuron* **12**, [1131-1138](#).
- Aslund et al., 1999. Aslund, F., Bessette, P.H., Georgiou, G., Beckwith, J., [1999](#). Efficient Production of Disulfide Bonded Proteins in the Cytoplasm in “Oxidizing” Mutants of *E. coli*. *in* *Novations* **10**, [11-12](#).
- Bailey et al., 1978. Bailey, G.S., Poulter, R.T.M., Stockwell, P.A., [1978](#). Gene duplication in tetraploid fish: Model for gene silencing at unlinked duplicated loci. *PNAS* **75**, [5575-5579](#).
- Belew & Porath, 1990. Belew, M., Porath, J., [1990](#). Immobilized metal ion affinity chromatography. Effect of solute structure, ligand density and salt concentration on the retention of peptide. *J. Chromatogr.* **516**, [333-354](#).
- Bereta & Bereta, 1995. Bereta, J., Bereta, M., [1995](#). Stimulation of glyceraldehyde-3-phosphate dehydrogenase mRNA levels by endogenous nitric oxide in cytokine-activated endothelium. *Biochem. Biophys. Res. Commun.* **217**, [363-369](#).
- Berget S., 1995. Berget, S., [1995](#). Exon recognition in vertebrate splicing. *J. Biol. Chem.* **270**, [2411-2414](#).
- Berglund et al., 1997. Berglund, J.A., Chua, K., Abovich, N., Reed, R., Rosbash, M., [1997](#). The splicing factor BBP interacts specifically with the pre-mRNA branchpoint sequence UACUAAC. *Cell* **89**, [781-787](#).
- Bern et al., 1991. Bern, H.A., McCormick, S.D., Kelley, K.M., Gray, E.S., Nishioka, R.S., Madsen, S.S., Tsai, P.I., [1991](#). Insulin-like growth factor “under water”: role in growth and function of fish and other

- poikilothermic vertebrates. In: Spencer, E.M. (ed.). *Modern Concepts of Insulin-like Growth Factors*. Elsevier Science Publishing Co., New York, NY, [85-96](#).
- Bessette et al., 1999. Bessette, P.H., Aslund, F., Beckwith, J., Georgiou, G., [1999](#). Efficient folding of proteins with multiple disulfide bonds in the *Escherichia coli* cytoplasm. *PNAS* **96**, [13703-13708](#).
- Bhatia et al., 1994. Bhatia, P., Taylor, W.R., Greenberg, A.H., Wright, J.A., [1994](#). Comparison of glyceraldehyde-3-phosphate dehydrogenase and 28S-ribosomal RNA gene expression as RNA loading controls for northern blot analysis of cell lines of varying malignant potential. *Anal. Biochem.* **216**, [223-226](#).
- Bjornsson et al., 1994. Bjornsson, B.T., Taranger, G.L., Hansen, T., Stefansson, S.O., Haux, C., [1994](#). The interrelationship between photoperiod, growth hormone and sexual maturation of adult Atlantic salmon (*Salmo salar*). *Gen. Comp. Endocrinol.* **93**, [70-81](#).
- Bolton et al., 1987. Bolton, J.P., Collie, N.L., Kawauchi, H., Hirano, T., [1987](#). Osmoregulatory actions of growth hormone in rainbow trout (*Salmo gairdneri*). *J. Endocrinol.* **112**, [63-68](#).
- Bornstein et al., 1988. Bornstein, P., McKay, J., Liska, D.J., Apone, S., Devarayalu, S., [1988](#). Interactions between the promoter and first intron are involved in transcriptional control of alpha 1(I) collagen gene expression. *Mol. Cell. Biol.* **8**, [4851-4857](#).
- Bowden & Georgiou 1990. Bowden, G.A., Georgiou, G., [1990](#). Folding and aggregation of α -lactamase in the periplasmic space of *Escherichia coli*. *J. Biol. Chem.* **265**, [16760-16766](#).
- Brodsky et al., 1992. Brodsky, L.I., Vasiliev, A.V., Kalaidzidis, Y.L., Osipov, Y.S., Tatuzov, R.L., Feranchuk, S.I., [1992](#). GeneBee: the program package for biopolymer structure analysis. *Dimacs* **8**, [127-139](#).
- Brodsky et al., 1995. Brodsky, L.I., Ivanov, V.V., Kalaidzidis, Y.L., Leontovich, A.M., Nikolaev, V.K., Feranchuk, S.I., Drachev, V.A., [1995](#). GeneBee-NET: Internet-based server for analyzing biopolymers structure. *Biochemistry* **60**, [923-928](#).
- Buratti & Baralle, 2004. Buratti, E., Baralle, F.E., [2004](#). Influence of RNA secondary structure on the Pre-mRNA splicing process. *Mol. Cell. Biol.* **24**, [10505-10514](#).
- Butler & Le Roith, 2001. Butler, A.A., Le Roith, D., [2001](#). Control of growth by the somatotrophic axis: growth hormone and insulin-like growth factors have related and independent roles. *Annual Rev. Physiol.* **63**, [141-164](#).
- Burzawa-Gerard & Delevallee-Fortier, 1992. Burzawa-Gerard, E., Delevallee-Fortier, B., [1992](#). Growth hormone implication during experimental induction of vitellogenesis by 17 β -estradiol in the female silver eel. *C.R. Aca. Sci. Paris, Ser. III* **314**, [411-416](#).
- Calladine & Drew, 1992. Calladine, C.R., Drew, H.R., [1992](#). Understanding DNA – the Molecule and how it works. *Academic Press*, San Diego, CA.
- Cao et al., 1989. Cao, Q.-P., Duguay, S.J., Plisetskaya, E., Steiner, D.F., Chan, S.J., [1989](#). Nucleotide sequence and growth hormone-regulated expression of salmon insulin-like growth factor I mRNA. *Mol. Endocrinol.* **3**, [2005-2010](#).
- Casas-Carrillo et al., 1997. Casas-Carrillo, E., Prill-Adams, A., Price, S.G., Clutter, A.C., Kirkpatrick, B.W., [1997](#). Relationship of growth hormone and insulin-like growth factor-I genotypes with growth and carcass traits in swine. *Animal Genetics* **28**, [88-93](#).

- Chalmers et al., 1990. Chalmers, J.J., Kim, E., Telford, J.N., Wong, E.Y., Tacon, W.C., Shuler, M.L., Wilson, D.B., [1990](#). Effects of temperature on *Escherichia coli* overproducing β -lactamase or human epidermal growth factor. *Appl. Environ. Microbiol.* **56**, [104-111](#).
- Chang Chang et al., 1998. Chang Chang, T.J., Juan, C.C., Yin, P.H., Chi, C.W., Tsay, H.J., [1998](#). Upregulation of β -actin, cyclophilin and GAPDH in NISI rat hepatoma. *Oncol. Rep.* **5**, [469-471](#).
- Chao et al., 1989. Chao, S.-C., Pan, F.-M., Chang, W.-C., [1989](#). Purification of carp growth hormone and cloning of the complementary DNA. *Biochim. Biophys. Acta* **1007**, [233-236](#).
- Chiou et al., 1990. Chiou, C.-S., Chen, H.-T., Chang, W.-C., [1990](#). The complete nucleotide sequence of the growth-hormone gene from the common carp (*Cyprinus carpio*). *Biochim. Biophys. Acta* **1087**, [91-94](#).
- Chrysogelos, 1993. Chrysogelos, S.A., [1993](#). Chromatin structure of the EGFR gene suggests a role for intron 1 sequences in its regulation in breast cancer cells. *Nucleic Acids Res.* **21**, [5736-5741](#).
- Collie & Stevens, 1985. Collie, N.L., Stevens, J.J., [1985](#). Hormonal effects on L-proline transport in coho salmon (*Oncorhynchus kisutch*) intestine. *Gen. Comp. Endocrinol.* **59**, [399-409](#).
- Collins-Racie et al., 1995. Collins-Racie, L.A., McColgan, J.M., Grant, K.L., DiBlasio-Smith, E.A., McCoy, J.M., LaVallie, E.R., [1995](#). Production of recombinant bovine enterokinase catalytic subunit in *Escherichia coli* using the novel secretory fusion partner DsbA. *Biotechnology* **13**, [982-987](#).
- Company et al., 1999. Company, R., Caldach-Giner, J.A., Kaushik, S., Perez-Sanchez, J., [1999](#). Growth performance and adiposity in gilthead sea bream (*Sparus auratus*): risks and benefits of high energy diets. *Aquaculture* **171**, [279-292](#).
- Cook & Peter, 1984. Cook, A.F., Peter, R.E., [1984](#). The effects of somatostatin on serum growth hormone levels in the goldfish, *Carassius auratus*. *Gen. Comp. Endocrinol.* **54**, [109-113](#).
- Cooke et al. 1997. Cooke, J., Nowak, M.A., Boerlijst, M., Maynard-Smith, J., [1997](#). Evolutionary origins and maintenance of redundant gene expression during metazoa development. *Trends Genet.* **13**, [360-364](#).
- Crothers et al. 1990. Crothers, D.M., Haran, T.E., Nadeau, J.G., [1990](#). Intrinsically bent DNA. *J. Biol. Chem.* **265**, [7093-7096](#).
- Czelusniak et al. 1982. Czelusniak, J., Goodman, M., Hewett-Emmett, D., Weiss, M.L., Venta, P.J., Tashian, R.E., [1982](#). Phylogenetic origins and adaptive evolution of avian and mammalian haemoglobin genes. *Nature* **298**, [297-300](#).
- David et al. 2003. David, L., Blum, S., Feldman, M.W., Lavi, U., Hillel, J., [2003](#). Recent duplication of the common carp (*Cyprinus carpio* L.) genome as revealed by analyses of microsatellite loci. *Mol. Biol. Evol.* **20**, [1425-1434](#).
- Davidson 1987. Davidson, M.B., [1987](#). Effect of growth hormone on carbohydrate and lipid metabolism. *Endocr. Rev.* **8**, [115-131](#).
- Davies & Hanyu 1986. Davies, P.R., Hanyu, I., [1986](#). Effect of temperature and photoperiod on sexual maturation and spawning of the common carp. I: under conditions of high temperature. *Aquaculture* **51**, [277-288](#).
- Davis et al. 1991. Davis, C.A., Homyard, D.P., Millen K.J., Joyner, A.L., [1991](#). Examining pattern formation in mouse, chicken and frog embryos with an En-specific antiserum. *Development* **2**, [287-298](#).
- de Jesus et al. 2002. de Jesus, E.G.T., Ayson, F.G., Amemiya, Y., Moriyama, S., Hyodo, S., Hirano, T., Kawauchi, H., [2002](#). Milkfish (*Chanos chanos*) growth hormone cDNA cloning and mRNA expression in embryos and early larval stages. *Aquaculture* **208**, [177-188](#).

- Deane & Woo 2006. Deane, E.E., Woo, N.Y., [2006](#). Molecular cloning of growth hormone from silver sea bream: effects of abiotic and biotic stress on transcriptional and translational expression. *Biochem. Biophys. Res. Commun.* **342**, [1077-1082](#).
- Degani & Gallaher 1985. Degani, G., Gallagher, M.L., [1985](#). Effects of dietary 17 α -methyltestosterone and bovine growth hormone on growth and food conversion of slow- and normally-growing American elvers (*Anguilla rostrata*). *Can. J. Fish. Aquat. Sci.* **42**, [185-189](#).
- Degani et al. 1996. Degani, G., Boker, K., Jackson, K., [1996](#). Growth hormone, gonad development, and steroid levels in female carp. *Comp. Biochem. Physiol.* **115C**, [133-140](#).
- Degani et al. 1998. Degani, G., Boker, K., Jackson, K., [1998](#). Growth hormone, sexual maturity and steroids in male carp. *Comp. Biochem. Physiol.* **120C**, [433-440](#).
- Delhase et al. 1993. Delhase, M., Vergani, P., Malur, A., Hooghe-Peters, E.L., Hooghe, R.J., [1993](#). The transcription factor Pit-1/GHF-1 is expressed in hemopoietic and lymphoid tissues. *Eur. J. Immunol.* **23**, [951-955](#).
- Devlin 1993. Devlin, R.H., [1993](#). Sequence of sockeye salmon type 1 and 2 growth hormone genes and the relationship of rainbow trout with Atlantic and Pacific salmon. *Can. J. Fish. Aquat. Sci.* **50**, [1738-1748](#).
- Diekmann & Wang 1985. Diekmann, S., Wang, J.C., [1985](#). On the sequence determinants and flexibility of the kinetoplast DNA fragment with abnormal gel electrophoretic mobilities. *J. Mol. Biol.* **186**, [1-11](#).
- Dolle et al. 1990. Dolle, P., Castrillo, J.L., Theill, L.E., Deerinck, T., Ellisman, M., Karin, M., [1990](#). Expression of GHF-1 protein in mouse pituitaries correlates both temporally and spatially with the onset of growth hormone gene activity. *Cell* **60**, [809-820](#).
- Donaldson et al. 1979. Donaldson, E.M., Fagerlund, U.H.M., Higgs, D.A., McBride, J.R., [1979](#). Hormonal enhancement of growth in fish. In: Hoare, W.S., Randall, D.J., Brett, J.R. (eds.). *Fish physiology*. Academic Press, New York, [455-597](#).
- Duan & Inui 1990a. Duan, C., Inui, Y., [1990a](#). Effects of recombinant eel growth hormone on the uptake of [³⁵S] sulfate by ceratobranchial cartilage of the Japanese eel, *Anguilla japonica*. *Gen. Comp. Endocrinol.* **79**, [320-325](#).
- Duan & Inui 1990b. Duan, C., Inui, Y., [1990b](#). Evidences for the presence of a somatomedin-like plasma factor(s) in the Japanese eel, *Anguilla japonica*. *Gen. Comp. Endocrinol.* **79**, [326-331](#).
- Duan & Hirano 1992. Duan, C., Hirano, T., [1992](#). Effects of insulin-like growth factor-I and insulin on the in vitro uptake of sulphate by eel branchial cartilage: evidence for the presence of independent hepatic and pancreatic sulphation factors. *J. Endocrinol.* **133**, [211-219](#).
- Duan et al. 1993. Duan, C., Duguay, S.J., Plisetskaya, E.M., [1993](#). Insulin-like growth factor I (IGF-I) mRNA expression in coho salmon, *Oncorhynchus kisutch*: tissue distribution and effects of growth hormone/prolactin family proteins. *Fish Physiol. Biochem.* **11**, [371-379](#).
- Duda & Palumbi 1999. Duda, T.F., Palumbi, S.R., [1999](#). Molecular genetics of ecological diversification: duplication and rapid evolution of toxin genes of the venomous gastropod. *PNAS* **96**, [6820-6823](#).
- Duret 2000. Duret, L., [2000](#). tRNA gene number and codon usage in the *C. elegans* genome are co-adapted for optimal translation of highly expressed genes. *Trends Genet.* **16**, [287-289](#).
- Endo et al. 1996. Endo, T., Ikeo, K., Gojobori, T., [1996](#). Large-scale search for genes on which positive selection may operate. *Mol. Biol. Evol.* **13**, [685-690](#).

- Engel et al. 1971. Engel, W., Faust, J., Wolf, U., [1971](#). Isoenzyme polymorphism of the sorbitol dehydrogenase and the NADP-dependent isocitrate dehydrogenases in the fish family Cyprinidae. *Anim. Blood Groups Biochem. Genet.* **2**, [127-133](#).
- FAO Fishery Statistics, [FAO Fishery Statistics 2005](#). available at www.fao.org/fi/statist/statist.asp
- Felsenstein 1985. Felsenstein, J., [1985](#). Confidence limits on phylogenies: An approach using bootstrap. *Evolution* **39**, [783-791](#).
- Ferris & Whitt 1977. Ferris, S., Whitt, G.S., [1977](#). The evolution of duplicate gene expression in the carp (*Cyprinus carpio*). *CMLS* **33**, [1299-1301](#).
- Ferris & Whitt 1979. Ferris, S.D., Whitt, G.S., [1979](#). Evolution of the differential regulation of duplicate genes after polyploidization. *J. Mol. Evol.* **12**, [267-317](#).
- Fisher 1935. Fisher, R.A., [1935](#). The sheltering of lethals. *Am. Naturalist* **69**, [446-445](#).
- Force et al. 1999. Force, A., Lynch, M., Pickett, F.B., Amores, A., Yan, Y., Postlethwait, J., [1999](#). Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* **151**, [1531-1545](#).
- Fostier et al. 1983. Fostier, A., Jalabert, B., Billard, R., Zohar, Y., [1983](#). The gonadal steroids. *Fish Physiology. Academic Press, New York*. [277-372](#).
- Fox et al. 2006. Fox, B.K., Riley, L.G., Hirano, T., Grau, E.G., [2006](#). Effects of fasting on growth hormone, growth hormone receptor, and insulin-like growth factor-I axis in seawater-acclimated tilapia, *Oreochromis mossambicus*. *Gen. Comp. Endocrinol.* **148**, [340-347](#).
- Franklin et al. 1991. Franklin, G.C., Donovan, M., Adam, G.I.R., Holmgren, L., Pfeifer-Ohlsson, S., Ohlsson, R., [1991](#). Expression of the human PDGF-B gene is regulated by both positively and negatively acting cell type-specific regulatory elements located in the first intron. *EMBO J.* **10**, [1365-1373](#).
- Friedman & Hughes 2001. Friedman, R., Hughes, A.L., [2001](#). Pattern and timing of gene duplication in animal genomes. *Genome Res.* **11**, [1842-1847](#).
- Friedman & Hughes 2003. Friedman, R., Hughes, A.L., [2003](#). The temporal distribution of gene duplication events in a set of highly conserved human gene families. *Mol. Biol. Evol.* **20**, [154-161](#).
- Froesch et al. 1985. Froesch, E.R., Schmid, C., Schwander, J., Zapf, J., [1985](#). Actions of insulin-like growth factors. *Annual Rev. Phys.* **47**, [443-467](#).
- Funkenstein & Cohen 1996. Funkenstein, B., Cohen, I., [1996](#). Ontogeny of growth hormone protein and mRNA in the gilthead sea bream *Sparus aurata*. *Growth Regul.* **6**, [16-21](#).
- Futami et al. 1997. Futami, J., Tsushima, Y., Murato, Y., Tada, H., Sasaki, J., Seno, M., Yamada, H., [1997](#). Tissue-specific expression of pancreatic-type RNase and RNase inhibitor in humans. *DNA Cell Biol.* **16**, [413-419](#).
- Futami et al. 2001. Futami, K., Komiya, T., Zhang, H., Okamoto, N., [2001](#). Differential expression of max and two types of *c-myc* genes in a tetraploid fish, the common carp (*Cyprinus carpio*). *Gene* **269**, [113-119](#).
- Futami et al. 2005. Futami, K., Zhang, H., Okamoto, N., [2005](#). Functional divergence of duplicated c-myc genes in a tetraploid fish, the common carp (*Cyprinus carpio*). *Gene* **363**, [61-66](#).
- Gabrielian & Pongor 1996. Gabrielian, A., Pongor, S., [1996](#). Correlation of intrinsic DNA curvature with DNA property periodicity. *FEBS Lett.* **393**, [65-68](#).
- Gabrielian et al. 1996. Gabrielian, A., Simoncsits, A., Pongor, S., [1996](#). Distribution of bending propensity in DNA sequences. *FEBS Lett.* **393**, [124-130](#).

- Galtier et al. 1996. Galtier, N., Gouy, M., Gautier, C., [1996](#). SEAVIEW and PHYLO_WIN: two graphic tools for sequence alignment and molecular phylogeny. *Comput. Appl. Biosci.* **12**, [543-548](#).
- Gardner & Barald 1992. Gardner, C.A., Barald, K.F., [1992](#). Expression patterns of engrailed-like proteins in the chick embryo. *Dev. Dyn.* **193**, [370-388](#).
- Gaut & Doebley 1997. Gaut, B.S., Doebley, J.F., [1997](#). DNA sequence evidence for the segmental allotetraploid origin of maize. *PNAS* **94**, [6809-6814](#).
- Gebhardt et al. 1999. Gebhardt, F., Zänker, K.S., Brandt, B., [1999](#). Modulation of epidermal growth factor receptor gene transcription by a polymorphic dinucleotide repeat in Intron 1. *J. Biol. Chem.* **19**, [13176-13180](#).
- Gibson & Spring 1998. Gibson, T.J., Spring, J., [1998](#). Genetic redundancy in vertebrates: polyploidy and persistence of genes encoding multidomain proteins. *PNAS* **14**, [46-49](#).
- Gohlke 1996-2007. Gohlke, C., [1996-2007](#). "DNA Curvature Analysis" software on the website <http://www.lfd.uci.edu/~gohlke/curve>.
- Gomez-Requeni et al. 2004. Gomez-Requeni, P., Mingarro, M., Calduch-Giner, J.A., Medale, F., Martin, S.A.M., Houlihan, D.F., Kaushik, S., Perez-Sanchez, J., [2004](#). Protein growth performance, amino acid utilization and somatotropic axis responsiveness to fish meal replacement by plant protein sources in gilthead sea bream (*Sparus aurata*). *Aquaculture* **232**, [493-510](#).
- Goodman 1981a. Goodman, M., [1981a](#). Decoding the pattern of protein evolution. *Prog. Biophys. Mol. Biol.* **38**, [105-164](#).
- Goodman 1981b. Goodman, M., [1981b](#). Globin evolution was apparently very rapid in early vertebrates: A reasonable case against the rate-constancy hypothesis. *J. Mol. Evol.* **17**, [114-120](#).
- Gray et al. 1990. Gray, E.S., Young, G., Bern, H.A., 1990. Radioreceptor assay for growth hormone in Coho salmon (*Oncorhynchus kisutch*) and its application to the study of stunting. *J. Exp. Zool.* **256**, [290-296](#).
- Grochowska et al. 2001. Grochowska, R., Sorensen, P., Zwierzchowski, L., Snochowski, M., Lovendahl, P., [2001](#). Genetic variation in stimulated GH release and in IGF-I of young dairy cattle and their associations with the leucine/valine polymorphism in the GH gene. *J. Anim. Sci.* **79**, [470-476](#).
- Gross & Nilsson 1999. Gross, R., Nilsson J., [1999](#). Restriction fragment length polymorphism at the growth hormone 1 gene in Atlantic salmon (*Salmo salar L.*) and its association with weight among the offspring of a hatchery stock. *Aquaculture* **173**, [73-80](#).
- Gross et al. 2002. Gross, R., Kohlmann, K., Kersten, P., [2002](#). PCR-RFLP analysis of the mitochondrial *ND-3/4* and *ND-5/6* gene polymorphisms in the European and East Asian subspecies of common carp (*Cyprinus carpio L.*). *Aquaculture* **204**, [507-516](#).
- Gu et al. 2002. Gu, Z., Nicolae, D., Lu, H.H.-S., Li, W.-H., [2002](#). Rapid divergence in expression between duplicate genes inferred from microarray data. *Trends Genet.* **18**, [609-613](#).
- Guigo et al. 1996. Guigo, R., Muchnik, I., Smith, T.F., [1996](#). Reconstruction of ancient molecular phylogeny. *Mol. Phyl. Evol.* **6**, [189-213](#).
- Hagerman 1990. Hagerman, P.J., [1990](#). Sequence-directed curvature of DNA. *Annual Rev. Biochem.* **59**, [755-781](#).
- Haldane 1933. Haldane, J.B.S., [1933](#). The part played by recurrent mutation in evolution. *Am. Naturalist* **67**, [5-19](#).

- Hall 1999. Hall, T.A., [1999](#). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl. Acids Symp. Ser.* **41**, [95-98](#).
- Harrington 1992. Harrington, R.E., 1992. DNA curving and bending in protein-DNA recognition. *Mol. Microbiol.* **6**, 2549-2555.
- Harrison et al. 2002. Harrison, P.M., Hegyi, H., Balasubramanian, S., Luscombe, N.M., Bertone, P., Echols, N., Johnson, T., Gerstein, M., [2002](#). Molecular fossils in the human genome: identification and analysis of the pseudogenes in chromosomes 21 and 22. *Genome Res.* **12**, [272-280](#).
- Harvey et al. 1995. Harvey, S., Scanes, C.G., Daughaday, W.H., [1995](#). Growth hormone. CRC Press, Boca Raton, FL.
- Harvey et al. 2004. Harvey, S., Baudet, M.L., Murphy, A., Luna, M., Hull, K.L., Aramburo, C., [2004](#). Testicular growth hormone (GH): GH expression in spermatogonia and primary spermatocytes. *Gen. Comp. Endocrinol.* **139**, [158-167](#).
- He & Zhang 2005. He, X., Zhang, J., [2005](#). Rapid subfunctionalization accompanied by prolonged and substantial neofunctionalization in duplicate gene evolution. *Genetics* **169**, [1157-1164](#).
- Hertel et al. 1997. Hertel, K.J., Lynch, K.W., Maniatis, T., [1997](#). Common themes in the function of transcription and splicing enhancers. *Curr. Opin. Cell Biol.* **9**, [350-357](#).
- Higg et al. 1975. Higg, D.A., Donaldson, E.M., Dye, H.M., McBride, J.R., [1975](#). A preliminary investigation of the effect of growth hormone on growth and muscle composition of coho salmon (*Oncorhynchus kisutch*). *Gen. Comp. Endocrinol.* **27**, [240-257](#).
- Hirano 1991. Hirano, T., [1991](#). Hepatic receptors for homologous growth hormone in the eel. *Gen. Comp. Endocrinol.* **81**, [383-390](#).
- Hochuli et al. 1987. Hochuli, E., Döbeli, H., Schacher, A., [1987](#). New metal chelate adsorbent selective for proteins and peptides containing neighbouring histidine residues. *J. Chromatogr.* **411**, [177-184](#).
- Hochuli et al. 1988. Hochuli, E., Bannwarth, W., Döbeli, R., Gentz, R., Stüber, D., [1988](#). Genetic approach to facilitate purification of recombinant proteins with a novel metal chelate adsorbent. *BioTechnology* **6**, [1321-1325](#).
- Høj et al. 1993. Høj, S., Fredholm, M., Larsen, N.J., Nielsen, V.H., [1993](#). Growth hormone gene polymorphism associated with selection for milk fat production in lines of cattle. *Animal Genetics* **24**, [91-96](#).
- Horant et al. 2003. Horant, M., Millar, D.S., Hedderich J., Lewis, G., Newsway, V., Mo, N., Fryklund, L., Procter, A.M., Krawczak, M., Cooper, D.N., [2003](#). Human growth hormone 1 (GH1) gene expression: complex haplotype-dependent influence of polymorphic variation in the proximal promoter and locus control region. *Human Mutation* **21**, [408-423](#).
- Hubbs 1955. Hubbs, C.L., [1955](#). Hybridization between fish species in nature. *Syst. Zool.* **4**, [1-20](#).
- Huggard & Habibi 1996. Huggard, D., Habibi, H.R., [1996](#). Effect of testosterone on growth hormone and maturational gonadotropin subunit gene expression in the cultured goldfish pituitary in vitro. In: Goetz, F. W., Thomas, P. (eds.). *Proceedings of the Fifth International Symposium on Reproduction Physiology of Fish*. Austin, TX, [19-21](#).
- Hughes 1994. Hughes, A.L., [1994](#). The evolution of functionally novel proteins after gene duplication. *Proc. R. Soc. Lond. Ser. B* **256**, [119-124](#).

- Hughes 1999. Hughes, A.L., [1999](#). [Adaptive Evolution of Genes and Genomes](#). Oxford University Press.
- Hughes & Nei 1989. Hughes, A.L., Nei, M., [1989](#). Evolution of the major histocompatibility complex: independent origin of nonclassical class I genes in different groups of mammals. *Mol. Biol. Evol.* **6**, [559-579](#).
- Hughes et al. 1990. Hughes, A.L., Ota, T., Nei, M., [1990](#). Positive Darwinian selection promotes charge profile diversity in the antigen-binding cleft of class I major-histocompatibility-complex molecules. *Mol. Biol. Evol.* **7**, [515-524](#).
- Hughes & Hughes 1993. Hughes, M.K., Hughes, A.L., [1993](#). Evolution of duplicate genes in a tetraploid animal, *Xenopus laevis*. *Mol. Biol. Evol.* **10**, [1360-1369](#).
- Hughes et al. 2000. Hughes, A.L., Green, J.A., Garbayo, J.M., Roberts, R.M., [2000](#). Adaptive diversification within a large family of recently duplicated placentally expressed genes. *PNAS* **97**, [3319-3323](#).
- Izadyar et al. 1999. Izadyar, F., Zhao, J., Van Tol, H.T.A., Colenbrander, B., Bevers, M.M., [1999](#). Messenger RNA expression and protein localization of growth hormone in bovine ovarian tissue and in cumulus oocyte complexes (COCs) during *in vitro* maturation. *Mol. Reprod. Develop.* **53**, [398-406](#).
- Jensen 1976. Jensen, R.A., [1976](#). Enzyme recruitment in the evolution of new function. *Annual Rev. Microbiol.* **30**, [409-425](#).
- Jeon et al. 1999. Jeon, J.-T., Carlborg, Ö., Törnsten, A., Giuffra, E., Amarger, V., Chardon, P., Andersson-Eklund, L., Andersson, K., Hansson, I., Lundström, K., Andersson, L., [1999](#). A paternally expressed QTL affecting skeletal and cardiac muscle mass in pigs maps to the IGF2 locus. *Nature Genet.* **21**, [157-158](#).
- Johansen et al. 1989. Johansen, B., Johnsen, O.C., Valla, S., [1989](#). The complete nucleotide sequence of the growth-hormone gene from Atlantic salmon (*Salmo salar*). *Gene* **77**, [317-324](#).
- Johnson et al. 2005. Johnson, A.D., Wang, D., Sadee, W., [2005](#). Polymorphisms affecting gene regulation and mRNA processing: Broad implications for pharmacogenetics. *Pharm. Therap.* **106**, [19-38](#).
- Joyner & Martin 1987. Joyner, A.L., Martin, G.R., [1987](#). En-1 and En-2, two mouse genes with sequence homology to the *Drosophila* engrailed gene: expression during embryogenesis. *Genes Dev.* **1**, [29-38](#).
- Jukes & Cantor 1969. Jukes, T.H., Cantor, C.R., [1969](#). Evolution of protein molecules. In: Munro, H.N. (ed.). *Mammalian protein metabolism*. Academic Press, New York, [121-132](#).
- Kashi et al. 1990. Kashi, Y., Tikochinsky, Y., Iraqi, F., Beckmann, J.S., Gruenbaum, Y., Soller, M., [1990](#). Large DNA fragments containing poly(TG) are highly polymorphic in a variety of vertebrates. *Nucleic Acids Res.* **18**, [1129-1132](#).
- Kashi et al. 1997. Kashi, Y., King, D., Soller, M., [1997](#). Simple sequence repeats as a source of quantitative genetic variation. *Trends Genet.* **13**, [74-78](#).
- Kato et al. 1992. Kato, G.J., Lee, W.M., Chen, L.L., Dang, C.V., [1992](#). Max: functional domains and interaction with c-Myc. *Genes Dev.* **6**, [81-92](#).
- Kausel et al. 1999. Kausel, G., Vera, M.I., San Martin, R., Figueroa, J., Molina, A., Muller, M., Martial, J., Krauskopf, M.J., [1999](#). Transcription factor pit-1 expression is modulated upon seasonal acclimatization of eurythermal ectotherms: identification of two pit-1 genes in the carp. *Cell. Biochem.* **75**, [598-609](#).

- Kiefhaber et al. 1991. Kiefhaber, T., Rudolph, R., Kohler, H.-H., Buchner, J., [1991](#). Protein aggregation *in vitro* and *in vivo*: A quantitative model of the kinetic competition between folding and aggregation. *BioTechnology* 9, [825-829](#).
- Kim et al. 1993. Kim, J.L., Nikolov, D.B., Burley, S.K., [1993](#). Crystal structure of a yeast TBP/TATA-box complex. *Nature* 365, [512-520](#).
- Kim & Burley 1994. Kim, J.L., Burley, S.K., [1994](#). A resolution refined structure of TBP recognizing the minor groove of TATAAAAG. *Nature Struct. Biol.* 1, [638-653](#).
- Kimura 1983. Kimura, M., [1983](#). The neutral theory of molecular evolution. Cambridge University Press, Cambridge.
- King 1994. King, D.G., [1994](#). Triple repeat DNA as a highly mutable regulatory system. *Science* 263, [595-596](#).
- King et al. 1997. King, D.G., Soller, M., Kashi, Y., [1997](#). Evolutionary tuning knobs. *Endeavour* 21, [36-40](#).
- King & Soller 1999. King, D.G., Soller, M., [1999](#). Variation and fidelity: the evolution of simple sequence repeats as functional elements in adjustable genes. In: Wasser, S.P. (ed.). *Evolutionary theory and processes: modern perspectives*. Kluwer Academic Publishers, Dordrecht, The Netherlands, [65-82](#).
- Kirby et al. 1995. Kirby, D.A., Muse, S.V., Stephan, W., [1995](#). Maintenance of pre-mRNA secondary structure by epistatic selection. *PNAS* 92, [9047-9051](#).
- Kirpichnikov 1981. Kirpichnikov, V.S., [1981](#). Genetic bases of fish selection. Springer Verlag, Berlin, New York.
- Klinkhardt et al. 1995. Klinkhardt, M., Tesche, M., Greven, H., 1995. Database of fish chromosomes. Westarp Wissenschaften.
- Kohlmann et al. 1997. Kohlmann, K., Kersten, P., Luczynski, M., [1997](#). New polymorphism in glucosephosphate isomerase (GPI) of a wild common carp (*Cyprinus carpio* L.) from Vietnam. *Animal Genetics* 28, [313](#).
- Kohlmann & Kersten 1999. Kohlmann, K., Kersten, P., [1999](#). Genetic variability of German and foreign common carp (*Cyprinus carpio* L.) populations. *Aquaculture* 173, [435-445](#).
- Kohlmann & Luczynski 2000. Kohlmann, K., Luczynski, M., [2000](#). Genetic control of glucosephosphate isomerase (GPI) in common carp (*Cyprinus carpio* L.). *J. Heredity* 91, [400-402](#).
- Kohlmann et al. 2003. Kohlmann, K., Gross, R., Murakaeva, A., Kersten, P., [2003](#). Genetic variability and structure of common carp (*Cyprinus carpio*) populations throughout the distribution range inferred from allozyme, microsatellite and mitochondrial DNA markers. *Aquatic Liv. Res.* 16, [421-431](#).
- Kopetzki et al. 1989. Kopetzki, E., Schumacher, G., Buckel, P., [1989](#). Control of formation of active soluble or inactive insoluble baker's yeast alpha-glucosidase PI in *Escherichia coli* by induction and growth conditions. *Mol. Gen. Genet.* 216, [149-155](#).
- Koren et al. 1989. Koren, Y., Sarid, S., Ber, R., Daniel, V., [1989](#). Carp growth hormone: molecular cloning and sequencing of cDNA. *Gene* 77, [309-315](#).
- Kumar et al. 2004. Kumar, S., Tamura, K., Nei, M., [2004](#). MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and Sequence Alignment. *Brief. Bioinform.* 5, [150-163](#).

- Künzler et al. 1995. Künzler, P., Matsuo, K., Schaffner, W., 1995. Pathological, physiological, and evolutionary aspects of short unstable DNA repeats in the human genome. *Biological Chemistry Hoppe-Seyler* 376, 201-211.
- Laemmli 1970. Laemmli, U.K., 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* 227, 680-685.
- Larhammar & Risinger 1994. Larhammar, D., Risinger, C., 1994. Molecular genetic aspects of tetraploidy in the common carp *Cyprinus carpio*. *Mol. Phylogenet. Evol.* 3, 59-68.
- Larhammar et al. 2002. Larhammar, D., Lundin, L.-G., Hallböök, F., 2002. The human Hox-bearing chromosome regions did arise by block or chromosome (or even genome) duplications. *Genome Res.* 12, 1910-1920.
- LaVallie et al. 1993. LaVallie, E.R., DiBlasio, E.A., Kovacic, S., Grant, K.L., Schendel, P.F., McCoy, J.M., 1993. A thioredoxin gene fusion expression system that circumvents inclusion body formation in the *E. coli* cytoplasm. *BioTechnology* 11, 187-193.
- Le Gac et al. 1992. Le Gac, F., Ollitraut, M., Loir, M., Bail, P.Y., 1992. Evidence for binding and action of growth hormone in trout testis. *Biol. Reprod.* 46, 949-957.
- Li et al. 1990. Li, S., Crenshaw, E.B., Rawson, E.J., Simmons, D.M., Swanson, L.W., Rosenfeld, M.G., 1990. Dwarf locus mutants lacking three pituitary cell types result from mutations in the POU-domain gene *pit-1*. *Nature* 347, 528-533.
- Li & Graur 1991. Li, W.-H., Graur, D., 1991. *Fundamentals of Molecular Evolution*. Sinauer, Sunderland, MA.
- Lin et al. 1993. Lin, X.-W., Lin, H.-R., Peter, R.E., 1993. Growth hormone and gonadotropin secretion in the common carp (*Cyprinus carpio* L.): In vitro interactions of gonadotropin-releasing hormone, somatostatin, and the dopamine agonist apomorphine. *Gen. Comp. Endocrinol.* 89, 62-71.
- Lloyd et al. 1993a. Lloyd, R.V., Jin, L., Kulig, E., Thiny, M.T., Fields, K., Landefeld, T.D., Camper, S.A., 1993a. Pit-1/GHF-1 transcription factor expression in rodent pituitaries. *Endocr. Pathol.* 4, 146-154.
- Lloyd et al. 1993b. Lloyd, R.V., Jin, L., Chandler, W.F., Horvath, E., Stefaneanu, L., Kovacs, K., 1993b. Pituitary specific transcription factor messenger ribonucleic expression in adenomatous and nontumorous human pituitary tissues. *Lab. Invest.* 69, 570-575.
- Lorens et al. 1989. Lorens, J., Nerland, A.H., Male, R., Lossius, I., Telle, W., Totland, G., 1989. The nucleotide sequence of Atlantic salmon growth hormone cDNA. *Nucleic Acids Res.* 17, 2352.
- Lundin 1993. Lundin, L.G., 1993. Evolution of the vertebrate genome as reflected in paralogous chromosomal regions in man and the house mouse. *Genomics* 16, 1-19.
- Lynch & Conery 2000. Lynch, M., Conery, J.S., 2000. The evolutionary fate and consequences of duplicate genes. *Science* 290, 1151-1155.
- Lynch & Force 2000. Lynch, M., Force, A., 2000. The probability of duplicate gene preservation by subfunctionalization. *Genetics* 154, 459-473.
- Lynch et al. 2001. Lynch, M., O'Hely, M., Walsh, B., Force, A., 2001. The probability of preservation of a newly arisen gene duplicate. *Genetics* 159, 1789-1804.
- Mannhalter et al. 2000. Mannhalter, C., Koizar, D., Mitterbauer, G., 2000. Evaluation of RNA isolation methods and reference genes for RT-PCR analyses of rare target RNA. *Clin. Chem. Lab. Med.* 38, 171-177.

- Marchant et al. 1987. Marchant, T.A., Fraser, R.A., Andrews, P.C., Peter, R.E., [1987](#). The influence of mammalian and teleost somatostatins on the secretion of growth hormone from goldfish (*Carassius auratus* L.) pituitary fragments in vitro. *Regul. Peptides* **17**, [41-52](#).
- Marchant et al. 1989. Marchant, T.A., Chang, J.P., Nahorniak, C.S., Peter, R.E., [1989](#). Evidence that gonadotropin-releasing hormone also functions as a growth hormone-releasing factor in the goldfish. *Endocrinology* **124**, [2509-2518](#).
- Marini et al. 1982. Marini, J.C., Levene, S.D., Crothers, D.M., Englung, P.T., [1982](#). Bent helical structure in kinetoplast DNA. *PNAS* **79**, [7664-7668](#).
- Matthews 1992. Matthews, K.S., [1992](#). DNA looping. *Microbiol. Rev.* **56**, [123-136](#).
- McCormick et al. 1992. McCormick, S.D., Kelley, K.M., Young, G., Nishioka, R.S., Bern, H.A., [1992](#). Stimulation of coho salmon growth by insulin-like growth factor I. *Gen. Comp. Endocrinol.* **86**, [398-406](#).
- McLysaght et al. 2002. McLysaght, A., Hokamp, K., Wolfe, K.H., [2002](#). Extensive genomic duplication during early chordate evolution. *Nature Genet.* **31**, [200-204](#).
- Meloni et al. 1998. Meloni, R., Albanese, V., Ravassard, P., Treilhou, F., Mallet, J., [1998](#). A tetranucleotide polymorphic microsatellite, located in the first intron of the tyrosine hydroxylase gene, acts as a transcription regulatory element *in vitro*. *Hum. Mol. Genet.* **7**, [423-428](#).
- Merritt & Quattro 2001. Merritt, T.J., Quattro, J.M., [2001](#). Evidence for a period of directional selection following gene duplication in a neurally expressed locus of triosephosphate isomerase. *Genetics* **159**, [689-697](#).
- Messier & Stewart 1997. Messier, W., Stewart, C-B., [1997](#). Episodic adaptive evolution of primate lysosomes. *Nature* **385**, [151-154](#).
- Meyer & Scharl 1999. Meyer, A., Scharl, M., [1999](#). Gene and genome duplications in vertebrates: the one-to-four (to-eight in fish) rule and the evolution of novel gene functions. *Curr. Opin. Cell Biol.* **11**, [699-704](#).
- Moav et al. 1975. Moav, R., Hulata, G., Wohlfarth, G., [1975](#). Genetic differences between the Chinese and European races of the common carp. 1. Analysis of genotype-environment interactions for growth rate. *Heredity* **34**, [323-340](#).
- Moav & Wohlfarth 1976. Moav, R., Wohlfarth, G.W., [1976](#). Two-way selection for growth rate in the common carp (*Cyprinus carpio* L.). *Genetics* **82**, [83-101](#).
- Moore 2000. Moore, M.J., [2000](#). Intron recognition comes of age. *Nat. Struct. Biol.* **7**, [14-16](#).
- Murakaeva et al. 2003. Murakaeva, A., Kohlmann, K., Kersten, P., Kamilov, B., Khabibullin, D., [2003](#). Genetic characterization of wild and domesticated common carp (*Cyprinus carpio* L.) populations from Uzbekistan. *Aquaculture* **218**, [153-166](#).
- Murphy et al. 1987. Murphy, L.J., Bell, G.I., Friesen, H.G., [1987](#). Growth hormone stimulates sequential induction of c-myc and insulin-like growth factor I expression *in vivo*. *Endocrinology* **120**, [1806-1812](#).
- Nadeau & Sankoff 1997. Nadeau, J.H., Sankoff, D., [1997](#). Comparable rates of gene loss and functional divergence after genome duplication early in vertebrate evolution. *Genetics* **147**, [1259-1266](#).
- Nashioka et al. 1988. Nashioka, R.S., Kelley, K.M., Bern, H.A., [1988](#). Control of prolactin and growth hormone secretion in teleost fishes. *Zool. Sci.* **5**, [267-280](#).

- Nei 1987. Nei, M., [1987](#). [Molecular evolutionary genetics](#). [Columbia University Press](#), [New York](#).
- Nei & Gojobori 1986. Nei, M., Gojobori, T., [1986](#). [Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitution](#). [Mol. Biol. Evol.](#) [3](#), [418-426](#).
- Nei & Kumar 2000. Nei, M., Kumar, S., [2000](#). [Molecular evolution and phylogenetics](#). [Oxford University Press](#), [New York](#).
- Nei et al. 2000. Nei, M., Rogozin, I.B., Piontkivska, H., [2000](#). [Purifying selection and birth-and-death evolution in the ubiquitin gene family](#). [PNAS](#) [97](#), [10866-10871](#).
- Nelson 1994. Nelson, J.S., [1994](#). [Fishes of the world](#). 3rd edition, [Wiley](#), [New York](#).
- Newcomb et al. 1997. Newcomb, R.D., Campbell, P.M., Ollis, D.L., Cheah, E., Russell, R.J., Oakeshott, J.G., [1997](#). [A single amino acid substitution converts a carboxylesterase to an organophosphorus hydrolase and confers insecticide resistance on a blowfly](#). [PNAS](#) [94](#), [7464-7468](#).
- Nezer et al. 1999. Nezer, C., Moreau, L., Brouwers, B., Coppieters, W., Detilleux, J., Hanset, R., Karim, L., Kvasz, A., Leroy, P., Georges, M., [1999](#). [An imprinted QTL with major effect on muscle mass and fat deposition maps to the IGF2 locus in pigs](#). [Nature Genet.](#) [21](#), [155-156](#).
- Nielsen et al. 1995. Nielsen, V.H., Larsen, N.J., Agergaard, N.A., [1995](#). [Association of DNA polymorphism in the growth-hormone gene with basal-plasma growth-hormone concentration and production traits in pigs](#). [J. Anim. Breed. Genet.](#) [112](#), [205-212](#).
- Nowak et al. 1997. Nowak, M.A., Boerlijst, M.C., Cooke, J., Smith, J.M., [1997](#). [Evolution of genetic redundancy](#). [Nature](#) [388](#), [167-171](#).
- Nygren et al. 1994. Nygren, P.-A., Stahl, S., Uhlen, M., [1994](#). [Engineering proteins to facilitate bioprocessing](#). [Trends Biotechnol.](#) [12](#), [184-188](#).
- Ohno 1970. Ohno, S., [1970](#). [Evolution by gene duplication](#). [George Allen & Unwin](#), [London](#).
- Ohno 1973. Ohno, S., [1973](#). [Ancient linkage groups and frozen accidents](#). [Nature](#) [244](#), [259-262](#).
- Ohno & Atkin 1966. Ohno, S., Atkin, N.B., [1966](#). [Comparative DNA values and chromosome complements of eight species of fishes](#). [Chromosoma](#) [18](#), [455-466](#).
- Ohno et al. 1967. Ohno, S., Muramoto, J., Christian, L., Atkin, N.B., [1967](#). [Diploid-tetraploid relationship among old-world members of the fish family Cyprinidae](#). [Chromosoma](#) [23](#), [1-19](#).
- Ohta 1994. Ohta, T., [1994](#). [Further examples of evolution by gene duplication revealed through DNA sequence comparisons](#). [Genetics](#) [138](#), [1331-1337](#).
- Ojima & Hitotsumachi 1967. Ojima, Y., Hitotsumachi, S., [1967](#). [Cytogenetic studies in lower vertebrates. IV. A note on the chromosomes of the carp \(*Cyprinus carpio*\) in comparison with those of the funa and the goldfish \(*Carassius auratus*\)](#). [Proc. Jpn. Acad.](#) [42](#), [163-167](#).
- Orgel 1977. Orgel, L.E., [1977](#). [Gene duplication and the origin of protein with novel functions](#). [J. Theor. Biol.](#) [67](#), [773](#).
- Orlando et al. 1998. Orlando, C., Pinzani, P., Pazzagli, M., [1998](#). [Developments in quantitative PCR](#). [Clin. Chem. Lab. Med.](#) [36](#), [255-269](#).
- Oster et al. 2002. Oster, S.K., Ho, C.S.W., Soucie, E.L., Penn, L.Z., [2002](#). [The myc Oncogene: Marvelously Complex](#). [Adv. Cancer Res.](#) [84](#), [81-154](#).

- Ota & Nei 1995. Ota, N., Nei, M., [1995](#). Evolution of immunoglobulin VH pseudogenes in chickens. *Mol. Biol. Evol.* **12**, [94-102](#).
- Otto & Whitton 2000. Otto, S.P., Whitton, J., [2000](#). Polyploid incidence and evolution. *Annual Rev. Genet.* **34**, [401-437](#).
- Perez-Martin et al. 1994. Perez-Martin, J., Rojo, F., de Lorenzo, V., [1994](#). Promoters responsive to DNA bending: a common theme in prokaryotic gene expression. *Microbiol. Rev.* **58**, [268-290](#).
- Perez-Sanchez et al. 1992. Perez-Sanchez, J., Weil, C., Lebail, P.Y., [1992](#). Effects of human insulin-like growth factor-I on release of growth hormone by rainbow trout (*Oncorhynchus mykiss*) pituitary cells. *J. Exp. Zool.* **262**, [287-290](#).
- Perez-Sanchez et al. 1995. Perez-Sanchez, J., Martf-Palanca, H., Kaushik, S., [1995](#). Ration size and protein intake affect circulating growth hormone concentration, hepatic growth hormone binding and plasma insulin-like growth factor-I immunoreactivity in a marine teleost, the gilthead sea bream (*Sparus aurata*). *J. Nutr.* **125**, [546-552](#).
- Perutz 1983. Perutz, M.F., [1983](#). Species adaptation in a protein molecule. *Mol. Biol. Evol.* **1**, [1-28](#).
- Peter 1986. Peter, R.E., [1986](#). Vertebrate neurohormonal systems. In: Pang, P.K.T., Schreibman, M.P. (eds.) *Vertebrate Endocrinology: Fundamentals and Biomedical Implications*. Vol 1. *Academic Press*, NY, pp. [57-102](#).
- Pfaffl 2001. Pfaffl, M., [2001](#). A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res.* **29**, [2002-2007](#).
- Pfaffe et al. 1992. Pfaffle, R.W., Dimattia, G.E., Parks, J.S., Brown, M.R., Wit, J.M., Jansen, M., Van der Nat, H., Van den Brande, J.L., Rosenfeld, M.G., Ingraham, H.A., [1992](#). Mutation of the POU-specific domain of Pit-1 and hypopituitarism without pituitary hypoplasia. *Science* **257**, [1118-1121](#).
- Piatigorsky & Wistow 1991. Piatigorsky, J., Wistow, G., [1991](#). The recruitment of crystallins: new functions precede gene duplication. *Science* **252**, [1078-1079](#).
- Porath et al. 1975. Porath, J., Carlsson, J., Olsson, I., Belfrage, G., [1975](#). Metal chelate affinity chromatography, a new approach to protein fractionation. *Nature* **258**, [598-599](#).
- Porath & Belew 1983. Porath, J., Belew, M., [1983](#). IMA-chromatography (immobilized ion affinity chromatography): Reflections of methodological development. In: Chaiken, I.M. (ed.). *Affinity chromatography and biological recognition: Proceedings of the 5th International Symposium on Affinity Chromatography and Biological Recognition*. *Academic Press*, Orlando, Florida, [173-189](#).
- Postlethwait et al. 1998. Postlethwait, J.H., Yan, Y.L., Gates, M.A., Horne, S., Amores, A., Brownlie, A., Donovan, A., Egan, E.S., Force, A., Gong, Z., [1998](#). Vertebrate genome evolution and the zebrafish gene map. *Nature Genet.* **18**, [345-349](#).
- Prinz et al. 1997. Prinz, W.A., Aslund, F., Holmgren, A., Beckwith, J., [1997](#). The Role of the Thioredoxin and Glutaredoxin Pathways in Reducing Protein Disulfide Bonds in the *Escherichia coli* Cytoplasm. *J. Biol. Chem.* **272**, [15661-15667](#).
- Raicu et al. 1972. Raicu, P., Taisescu, E., Cristian, A., [1972](#). Diploid chromosome complement of the carp. *Cytologia* **37**, [355-358](#).
- Raymond & Rousset 1995. Raymond, M., Rousset, F., [1995](#). GENEPOP (Version 1.2): Population Genetics Software for Exact Tests and Ecumenicism. *J. Heredity* **86**, [248-249](#).

- Renaville et al. 2002. Renaville, R., Hammadi, M., Portetelle, D., 2002. Role of the somatotrophic axis in the mammalian metabolism. *Domest. Anim. Endocrinol.* 23, 351-360.
- Rhodes 1979. Rhodes, D., 1979. Nucleosome cores reconstituted from poly (dA-dT) and the octamer of histones. *Nucleic Acids Res.* 6, 1805-1816.
- Risinger & Larhammar 1993. Risinger, C., Larhammar, D., 1993. Multiple loci for synapse protein SNAP-25 in the tetraploid goldfish. *PNAS* 268, 24408-24414.
- Rosatto et al. 1999. Rosatto, N., Pontremoli, R., De Ferrari, G., Ravazzolo, R., 1999. Intron 16 insertion of the angiotensin converting enzyme gene and transcriptional regulation. *Nephrol. Dialys. Transplan.* 14, 868-871.
- Rosenberg 1995. Rosenberg, H.F., 1995. Recombinant human eosinophil cationic protein. Ribonuclease activity is not essential for cytotoxicity. *J. Biol. Chem.* 270, 7876-7881.
- Rouquier et al. 2000. Rouquier, S., Blancher, A., Giorgi, D., 2000. The olfactory receptor gene repertoire in primates and mouse: evidence for reduction of the functional fraction in primates. *PNAS* 97, 2870-2874.
- Roy et al. 1993. Roy, A.L., Carruthers, C., Gutjahr, T., Roeder, R.G., 1993. Direct role for Myc in transcription initiation mediated by interactions with TFII-I. *Nature* 365, 359-361.
- Rozen & Skaletsky 2000. Rozen, S., Skaletsky, H., 2000. Primer3 on the WWW for general users and for biologist programmers. *Methods Mol. Biol.* 132, 365-386.
- Rubin et al. 2000. Rubin, G.M., Yandell, M.D., Wortman, J.R., Miklos, G.L.G., Nelson, C.R., Hariharan, I.K., Fortini, M.E., Li, P.W., Apweiler, R., Fleischmann, W., 2000. Comparative genomics of the eukaryotes. *Science* 287, 2204-2215.
- Rudolph 1996. Rudolph, R., 1996. Successful protein folding on an industrial scale. In: Cleland, J.L., Craik, C.S. (eds.). *Protein Engineering: Principles and Practice*. John Wiley and Sons, New York, 283-298.
- Saitou & Nei 1987. Saitou, N., Nei, M., 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* 4, 406-425.
- Sakamoto & Hirano 1991. Sakamoto, T., Hirano, T., 1991. Growth hormone receptors in the liver and osmoregulatory organs of rainbow trout: characterization and dynamics during adaptation to seawater. *J. Endocrinol.* 130, 425-433.
- Sarmiento et al. 2000. Sarmiento, J., Leal, S., Quezada, C., Kausel, G., Figueroa, J., Vera, M. I., Krauskopf, M., 2000. Environmental acclimatization of the carp modulates the transcription of β -actin. *J. Cell. Biochem.* 80, 223-228.
- Schein 1989. Schein, C.H., 1989. Production of soluble recombinant proteins in bacteria. *BioTechnology* 7, 1141-1148.
- Schein & Noteborn 1988. Schein, C.H., Noteborn, M.H.M., 1988. Formation of soluble recombinant proteins in *Escherichia coli* is favored by lower growth temperature. *BioTechnology* 6, 291-294.
- Schmid et al. 1999a. Schmid, A., Näf, E., Kloas, W., Reinecke, M., 1999a. IGF-I and IGF-II in the ovary of the bony fish, *Oreochromis mossambicus*, the tilapia: In situ hybridisation, immunohistochemical localisation, northern blot and cDNA sequences. *Mol. Cell. Endocrinol.* 156, 141-149.

- Schmid et al. 1999b. Schmid, A., Kloas, W., Reinecke, M., [1999b](#). Establishment of a primary liver cell culture from a teleost, *Oreochromis mossambicus*, the tilapia: A valid tool for physiological studies. *Proceedings of the 16th ESACT Meeting*. Lugano, Switzerland, [April 25th to 29th 1999](#), [1-3](#).
- Schmid et al. 2000. Schmid, A., Reinecke, M., Kloas, W., [2000](#). Primary cultured hepatocytes of the bony fish, *Oreochromis mossambicus*, the tilapia: A valid tool for physiological studies on insulin-like growth factor-I (IGF-I) expression in liver. *J. Endocrinol.* **166**, [265-273](#).
- Schmid et al. 2003. Schmid, A., Lutz, I., Kloas, W., Reinecke, M., [2003](#). Thyroid hormone stimulates hepatic IGF-I mRNA expression in a bony fish, the tilapia *Oreochromis mossambicus*, *in vitro and in vivo*. *Gen. Comp. Endocrinol.* **130**, [129-134](#).
- Sekine et al. 1989. Sekine, S., Mizukami, T., Saito, A., Kawauchi, H., Itoh, S., [1989](#). Isolation and characterization of a novel growth hormone cDNA from chum salmon (*Oncorhynchus keta*). *Biochim. Biophys. Acta* **1009**, [117-120](#).
- Shakke et al. 1994. Shakke, Z., Guzikevich-Guerstein, G., Frolow, F., Rabinovich, D., Joachimiak, A., Sigler, P.B., [1994](#). Determinants of repressor/operator recognition from the structure of the trp operator binding site. *Nature* **368**, [469-473](#).
- Shamblott et al. 1995. Shamblott, M.J., Cheng, C.M., Bolt, D., Chen, T.T., [1995](#). Appearance of insulin-like growth factor mRNA in the liver and pyloric ceca of a teleost in response to exogenous growth hormone. *PNAS* **92**, [6943-6946](#).
- Shatzman 1990. Shatzman, A.R., [1990](#). Gene expression using gram-negative bacteria. *Curr. Opin. Biotechnol.* **1**, [5-11](#).
- Shih et al. 2002. Shih, Y.P., Kung, W.M., Chen, J.C., Yeh, J.H., Wang, A.H.J., Wang, T.F., [2002](#). High throughput screening of soluble recombinant proteins. *Protein Sci.* **11**, [1714-1719](#).
- Sica et al. 1992. Sica, A., Tan, T.H., Rice, N., Kretschmar, M., Gosh, P., Young, H.A., [1992](#). The c-Rel protooncogene product c-Rel but not NF-kappa B binds to the intronic region of the human interferon-gamma gene at a site related to an interferon-stimulable response element. *PNAS* **89**, [1740-1744](#).
- Simpson & Kunzler 1979. Simpson, R.T., Kunzler, P., [1979](#). Chromatin and core particles formed from the inner histones and synthetic polydeoxyribonucleotides of defined sequence. *Nucleic Acids Res.* **33**, [1387-1415](#).
- Singh et al. 1988. Singh, H., Griffith, R.W., Takahashi, A.P., Kawauchi, H., Stegeman, J.J., [1988](#). Regulation of gonadal steroidogenesis in *Fundulus heteroclitus* by recombinant salmon growth hormone and purified salmon prolactin. *Gen. Comp. Endocrinol.* **72**, [144-153](#).
- Singh & Thomas 1993. Singh, H., Thomas, P., [1993](#). Mechanism of stimulatory action of growth hormone on ovarian steroidogenesis in spotted seatrout (*Cynoscion nebulosus*). *Gen. Comp. Endocrinol.* **89**, [341-353](#).
- Smith & Valcárcel 2000. Smith, C.W., Valcárcel, J., [2000](#). Alternative pre-mRNA splicing: the logic of combinatorial control. *Trends Biochem. Sci.* **25**, [381-388](#).
- Smith & Pidgeon 1986. Smith, M.C., Pidgeon, C., [1986](#). Process for purifying proteins and compounds useful in such process. *U.S. Patent No. 4,569,794*.

- Smith et al. 1987. Smith, M.C., Furman, T.C., Pidgeon, C., [1987](#). Immobilized iminodiacetic acid metal peptide complexes. Identification of chelating peptide purification handles for recombinant proteins. *Inorg. Chem.* **26**, [1965-1969](#).
- Smith et al. 1988. Smith, M.C., Furman, T.C., Ingolia, T.D., Pidgeon, C., [1988](#). Chelating-peptide immobilized metal ion affinity chromatography. A new concept in affinity chromatography for recombinant proteins. *J. Biol. Chem.* **263**, [7211-7215](#).
- Spring 1997. Spring, J., [1997](#). Vertebrate evolution by interspecific hybridization - are we polyploid? *FEBS Lett.* **400**, [2-8](#).
- Stacey et al. 1984. Stacey, N.E., Mackenzie, D.S., Marchant, T.A., Kyle, A.L., Peter, R.E., [1984](#). Endocrine changes during natural spawning in the white sucker *Catostomus commersoni*. *Gen. Comp. Endocrinol.* **56**, [333-348](#).
- Starr et al. 1995. Starr, D.B., Hoopes, B.C., Hawley, D.K., [1995](#). DNA bending is an important component of site-specific recognition by the TATA binding protein. *J. Mol. Biol.* **250**, [434-446](#).
- Steffens 1980. Steffens, W., [1980](#). *Der Karpfen, Cyprinus carpio*. 5th edition, A. Ziemsen Verlag, Wittenberg Lutherstadt.
- Stewart et al. 1998. Stewart, E.J., Aslund, F., Beckwith, J., [1998](#). Disulfide bond formation in the *Escherichia coli* cytoplasm: an in vivo role reversal of the thioredoxins. *EMBO J.* **17**, [5543-5550](#).
- Strandberg & Enfors 1991. Strandberg, L., Enfors, S.O., [1991](#). Factors influencing inclusion body formation in the production of a fused protein in *Escherichia coli*. *Appl. Environ. Microbiol.* **57**, [1669-1674](#).
- Sumpter et al. 1991. Sumpter, J.R., Lincoln, R.F., Bye, V.J., Carragher, J.F., Le Bail, P.Y., [1991](#). Plasma growth hormone levels during sexual maturation in diploid and triploid rainbow trout (*Oncorhynchus mykiss*). *Gen. Comp. Endocrinol.* **83**, [103-110](#).
- Sun & Farmanfarmaian 1992. Sun, L.-Z., Farmanfarmaian, A., [1992](#). Biphasic action of growth hormone on intestinal amino acid absorption in striped bass hybrids. *Comp. Biochem. Physiol.* **103A**, [381-390](#).
- Suzuki & Yagi 1995. Suzuki, M., Yagi, N., [1995](#). Stereochemical basis of DNA bending by transcription factors. *Nucleic Acids Res.* **23**, [2083-2091](#).
- Tajima 1993. Tajima, F., [1993](#). Simple methods for testing the molecular evolutionary clock hypothesis. *Genetics* **135**, [599-607](#).
- Tajima & Nei 1984. Tajima, F., Nei, M., [1984](#). Estimation of evolutionary distance between nucleotide sequences. *Mol. Biol. Evol.* **1**, [269-285](#).
- Takahata & Maruyama 1979. Takahata, N., Maruyama, T., [1979](#). Polymorphism and loss of duplicate gene expression: A theoretical study with application to tetraploid fish. *PNAS* **76**, [4521-4525](#).
- Tamura & Nei 1993. Tamura, K., Nei, M., [1993](#). Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol. Biol. Evol.* **10**, [512-526](#).
- Tanaka & Nei 1989. Tanaka, T., Nei, M., [1989](#). Positive Darwinian selection observed at the variable-region genes of immunoglobulins. *Mol. Biol. Evol.* **6**, [447-459](#).
- Taylor et al. 2001. Taylor, J.S., Van de Peer, Y., Braasch, I., Meyer, A., [2001](#). Comparative genomics provides evidence for an ancient genome duplication event in fish. *Phil. Trans. R. Soc. Lond. B* **356**, [1661-1679](#).

- Thompson et al. 1994. Thompson, J.D., Higgins, D.G., Gibson, T.J., [1994](#). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, [4673-4680](#).
- Travers & Klug 1990. Travers, A.A., Klug, A., [1990](#). Bending of DNA in nucleoprotein complexes. In: Cozzarelli, N.R., Wang, J.C. (eds.). *DNA Topology and its Biological effects*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, [57-106](#).
- Trifonov 1985. Trifonov, E.N., [1985](#). Curved DNA. *CRC Crit. Rev. Biochem.* **19**, [89-106](#).
- Trudeau et al. 1992. Trudeau, V.L., Somoza, G.M., Nahorniak, C.S., Peter, R.E., [1992](#). Interactions of estradiol with gonadotropin-releasing hormone and thyrotropin-releasing hormone in the control of growth hormone secretion in the goldfish. *Neuroendocrinology* **56**, [483-490](#).
- Uyeno & Smith 1972. Uyeno, T., Smith, G.R., [1972](#). Tetraploid origin of the karyotype of catostomid fishes. *Science* **175**, [644-646](#).
- Van der Peer et al. 2001. Van de Peer, Y., Taylor, J.S., Braasch, I., Meyer, A., [2001](#). The ghost of selection past: rates of evolution and functional divergence of anciently duplicated genes. *J. Mol. Evol.* **53**, [436-446](#).
- Van der Kraak et al. 1990. Van der Kraak, G., Rosenblum, P.M., Peter, R.E., [1990](#). Growth hormone dependent potentiation of gonadotropin-stimulated steroid production by ovarian follicles of the goldfish. *Gen. Comp. Endocrinol.* **79**, [233-239](#).
- Van Laere et al. 2003. Van Laere, A.-S., Nguyen, M., Braunschweig, M., Nezer, C., Collette, C., Moreau, L., Archibald, A.L., Haley, C.S., Buys, N., Tally, M., Andersson, G., Georges, M., Andersson, L., [2003](#). A regulatory mutation in IGF2 causes a major QTL effect on muscle growth in the pig. *Nature* **425**, [832-836](#).
- Vaughan et al. 1992. Vaughan, J.M., Rivier, J., Spiess, J., Peng, C., Chang, J.P., Peter, R.E., Vale, W., [1992](#). Isolation and characterization of hypothalamic growth-hormone releasing factor from common carp, *Cyprinus carpio*. *Neuroendocrinology* **56**, [539-549](#).
- Vision et al. 2000. Vision, T.J., Brown, D.G., Tanksley, S.D., [2000](#). The origins of genomic duplications in *Arabidopsis*. *Science* **290**, [2114-2117](#).
- Vong et al. 2003. Vong, Q.P., Chan, K.M., Cheng, C.H.K., [2003](#). Quantification of common carp (*Cyprinus carpio*) IGF-I and IGF-II mRNA by real-time PCR: differential regulation of expression by GH. *J. Endocrinol.* **178**, [513-521](#).
- Voss & Rosenfeld 1992. Voss, J.W., Rosenfeld, M.G., [1992](#). Anterior pituitary development: Short tales from dwarf mice. *Cell* **70**, [527-530](#).
- Wagner 2000. Wagner, A., [2000](#). Decoupled evolution of coding region and mRNA expression pattern after gene duplication: implications for the neutralist-selection debate. *PNAS* **97**, [6579-6584](#).
- Wallis 1994. Wallis, M., [1994](#). Variable evolutionary rates in the molecular evolution of mammalian growth hormone. *J. Mol. Evol.* **38**, [619-627](#).
- Wallis 1996. Wallis, M., [1996](#). The molecular evolution of vertebrate growth hormones: A pattern of near-stasis interrupted by sustained bursts of rapid change. *J. Mol. Evol.* **43**, [93-100](#).
- Walsh 1995. Walsh, J.B., [1995](#). How often do duplicated genes evolve new functions? *Genetics* **139**, [421-428](#).
- Wang et al. 1989. Wang, A.M., Doyle, M.V., Mark, D.F., [1989](#). Quantitation of mRNA by the polymerase chain reaction. *PNAS* **86**, [9717-9721](#).

- Watterson 1983. Watterson, G.A., [1983](#). On the time for gene silencing at duplicate loci. *Genetics* **105**, [745-766](#).
- Wattier et al. 1998. Wattier, R., Engel, C.R., Saumitou-Laprade, P., Valero, M., [1998](#). Short allele dominance as a source of heterozygote deficiency at microsatellite loci: experimental evidence at the dinucleotide locus Gv1CT in *Gracilaria gracilis* (Rhodophyta). *Mol. Ecol.* **7**, [1569-1573](#).
- Wendel 2000. Wendel, J.F., [2000](#). Genome evolution in polyploids. *Plant Mol. Biol.* **42**, [225-249](#).
- Werth & Windham 1991. Werth, C., Windham, M., [1991](#). A model for divergent, allopatric speciation of polyploid pteridophytes resulting from silencing of duplicate-gene expression. *Am. Naturalist* **137**, [515-526](#).
- White et al. 1992. White, B.M., Carvalho, M., Derse, D., O'Brien, S.J., Dean, M., [1992](#). Detecting single base substitutions as heteroduplex polymorphisms. *Genomics* **12**, [301-306](#).
- Wittbrodt et al. 1998. Wittbrodt, J., Meyer, A., Schartl, M., [1998](#). More genes in fish? *BioEssays* **20**, [511-512](#).
- Wittwer et al. 1997. Wittwer, C.T., Ririe, K.M., Andrew, R.V., David, D.A., Gundry, R.A., Balis, U.J., [1997](#). The LightCycler: a microvolume multisample fluorimeter with rapid temperature control. *Biotechniques* **22**, [176-181](#).
- Wohlfarth 1986. Wohlfarth, G.W., [1986](#). Selective breeding of the common carp. In: Billard R., Marcel, J. (eds.) *Aquaculture of cyprinids*. INRA, Jouy-en-Josas, France, [195-208](#).
- Wohlfarth et al. 1964. Wohlfarth, G.W., Lahmann, M., Moav, R., [1964](#). Genetic improvement of carp: V. A comparison of fall and winter growth of several carp progeny. *Bamidgeh* **16**, [71-76](#).
- Wohlfarth et al. 1975. Wohlfarth, G.W., Moav, R., Hulata, G., [1975](#). Genetic differences between the Chinese and European races of the common carp. 2. Multi-character variation – a response to the diverse methods of fish cultivation in Europe and China. *Heredity* **34**, [341-350](#).
- Wolf et al. 1969. Wolf, U., Ritter, H., Atkin, N.B., Ohno, S., [1969](#). Polyploidization in the fish family Cyprinidae, Order Cypriniformes. *Humangenetik* **7**, [240-244](#).
- Wolfe & Shields 1997. Wolfe, K.H., Shields, D.C., [1997](#). Molecular evidence for an ancient duplication of the entire yeast genome. *Nature* **387**, [708-713](#).
- Wong et al. 1998. Wong, L., Pearson, H., Fletcher, A., Marquis, C.P., Mahler, S., [1998](#). Comparison of the efficiency of M-MuLV reverse transcriptase, RNase M-MuLV reverse transcriptase and AMV reverse transcriptase for the amplification of human immunoglobulin genes. *Biotechnol. Techniques* **12**, [485-489](#).
- Woods & Buth 1984. Woods, T.D., Buth, D.G., [1984](#). High levels of gene silencing in the tetraploid goldfish. *Biochem. Syst. Ecol.* **12**, [415-421](#).
- Wu & Li 1985. Wu, C.I., Li, W.H., [1985](#). Evidence for higher rates of nucleotide substitution in rodents than in man. *PNAS* **82**, [1741-1745](#).
- Yada et al. 2005. Yada, T., Muto, K., Azuma, T., Hyodo, S., Schreck, C.B., [2005](#). Cortisol stimulates growth hormone gene expression in rainbow trout leucocytes in vitro. *Gen. Comp. Endocrinol.* **142**, [248-255](#).
- Yamada et al. 1993. Yamada, S., Hata, J., Yamashita, S.J., [1993](#). Molecular cloning of fish Pit-1 cDNA and its functional binding to promoter of gene expressed in the pituitary. *Biol. Chem.* **268**, [24361-24366](#).
- Yamashita et al. 1986. Yamashita, S., Slanina, S., Kado, H.S., Melmed, S., [1986](#). Autoregulation of pituitary growth hormone messenger ribonucleic acid levels in rats bearing transplantable mammosomatotrophic pituitary tumors. *Endocrinology* **118**, [915-918](#).

- Yang et al. 1997. Yang, B.Y., Chan, K.M., Lin, C.M., Chen, T.T., [1997](#). Characterization of rainbow trout (*Oncorhynchus mykiss*) growth hormone 1 gene and the promoter region of growth hormone 2 gene. *Archive Biochem. Biophys.* **340**, [2359-2368](#).
- Yang et al. 1999. Yang, B.Y., Greene, M.W., Chen, T.T., [1999](#). Early embryonic expression of the growth hormone family protein genes in the developing rainbow trout, *Oncorhynchus mykiss*. *Mol. Reprod. Dev.* **53**, [127-134](#).
- Yao et al. 1991. Yao, K., Niu, P., LeGac, F., Le Bail, P.-Y., [1991](#). Presence of specific growth hormone binding sites in rainbow trout (*Oncorhynchus mykiss*) tissues: Characterization of the hepatic receptor. *Gen. Comp. Endocrinol.* **81**, [72-82](#).
- Yokoyama & Youkoyama 1989. Yokoyama, S., Yokoyama, R., [1989](#). Molecular evolution of human visual pigment genes. *Mol. Biol. Evol.* **6**, [186-197](#).
- Young et al. 1983. Young, G., Crim, L.W., Kagawa, H., Kambegawa, A., Nagahama, Y., [1983](#). Plasma $17\alpha,20\beta$ -dihydroxy-4-pregnen-3-one levels during sexual maturation of amago salmon (*Oncorhynchus rhodurus*): Correlations with plasma gonadotropin and in vitro production by ovary follicles. *Gen. Comp. Endocrinol.* **51**, [69-105](#).
- Zhang 2003. Zhang, J., [2003](#). Evolution by gene duplication: an update. *Trends Ecol. Evol.* **18**, [292-298](#).
- Zhang & Snyder 1992. Zhang, J., Snyder, S.H., [1992](#). Nitric oxide stimulates auto-ADP-ribosylation of glyceraldehyde 3 phosphate dehydrogenase. *PNAS* **89**, [9382-9385](#).
- Zhang et al. 1998. Zhang, J., Rosenberg, H.F., Nei, M., [1998](#). Positive Darwinian selection after gene duplication in primate ribonuclease genes. *PNAS* **95**, [3708-3713](#).
- Zhang et al. 2000. Zhang, J., Dyer, K.D., Rosenberg, H.F., [2000](#). Evolution of the rodent eosinophil-associated RNase gene family by rapid gene sorting and positive selection. *PNAS* **97**, [4701-4706](#).
- Zhang et al. 2002. Zhang, J., Zhang, Y., Rosenberg, H.F., [2002](#). Adaptive evolution of a duplicated pancreatic ribonuclease gene in leaf-eating monkey. *Nature Genet.* **30**, [411-415](#).
- Zheng et al. 2003. Zheng, L., Baumann, U., Reymond, J.L., [2003](#). Production of a functional catalytic antibody ScFv–NusA fusion protein in bacterial cytoplasm. *J. Biochem.* **133**, [577-581](#).
- Zuckerandl, 2001. Zuckerandl, E., [2001](#). Intrinsically driven changes in gene interaction complexity. I. The spontaneous growth of regulatory complexes. *J. Mol. Evol.* **53**, [539-554](#).

Appendix

Alignment A. 1 Alignment of fragments obtained with primer pair GH-a + GH-b: *GH I* from first exon to 3' non-coding end. EU - sample from Europe (River Rhine), clones 1 and 2; CA I - clones 1 and 2, CA II - clones 1 and 2, two samples from Central Asia (Uzbekistan); EA - sample from East Asia (China), two identical clones. Koren - cDNA from Koren et al. (1989) (*GH I*); Gene 1990 - complete gene sequence from Chiou et al. (1990) (*GH I*); Chao - cDNA from Chao et al. (1989) (*GH II*)

```

      *      20      *      40      *      60      *      80      *      100
EU1 : CTAAACAAATCTAAATATTCCTTATCTTTCTTTCATTTATTTTACACCTTAACAGATTTCATCTTATTTTACATATCCCTATTTA : 107
EU2 : CTAAACAAATCTAAATATTCCTTATCTTTCTTTCATTTATTTTACACCTTAACAGATTTCATCTTATTTTACATATCCCTATTTA : 107
CAI1 : CTAAACAAATCTAAATATTCCTTATCTTTCTTTCATTTATTTTACACCTTAACAGATTTCATCTTATTTTACATATCCCTATTTA : 107
CAI2 : CTAAACAAATCTAAATATTCCTTATCTTTCTTTCATTTATTTTACACCTTAACAGATTTCATCTTATTTTACATATCCCTATTTA : 107
CAII1 : CTAAACAAATCTAAATATTCCTTATCTTTCTTTCATTTATTTTACACCTTAACAGATTTCATCTTATTTTACATATCCCTATTTA : 107
CAII2 : CTAAACAAATCTAAATATTCCTTATCTTTCTTTCATTTATTTTACACCTTAACAGATTTCATCTTATTTTACATATCCCTATTTA : 107
EA : CTAAACAAATCTAAATATTCCTTATCTTTCTTTCATTTATTTTACACCTTAACAGATTTCATCTTATTTTACATATCCCTATTTA : 107
Koren : CTAAACAAATCTAAATATTCCTTATCTTTCTTTCATTTATTTTACACCTTAACAGATTTCATCTTATTTTACATATCCCTATTTA : 19
Gene1990 : CTAAACAAATCTAAATATTCCTTATCTTTCTTTCATTTATTTTACACCTTAACAGATTTCATCTTATTTTACATATCCCTATTTA : 107
Chao : CTAAACAAATCTAAATATTCCTTATCTTTCTTTCATTTATTTTACACCTTAACAGATTTCATCTTATTTTACATATCCCTATTTA : 18
      cTGAGCGAAATGGCTAGGtatgtctgtgatgctttctgttgcagt attgtgacacttaggacatggatttcattctttttgacatatccctatttagtta

      *      120      *      140      *      160      *      180      *      200
EU1 : TTTATTAATACATATTATATATTTTATTACCTTAAAAAATAATTTTATAATTAAATAAAAAAAAAAAAAATATATAAAAAAAATATCCAAAA : 213
EU2 : TTTATTAATACATATTATATATTTTATTACCTTAAAAAATAATTTTATAATTAAATAAAAAAAAAAAAAATATATAAAAAAAATATCCAAAA : 212
CAI1 : TTTATTAATACATATTATATATTTTATTACCTTAAAAAATAATTTTATAATTAAATAAAAAAAAAAAAAATATAAAAAAAATATCCAAAA : 212
CAI2 : TTTATTAATACATATTATATATTTTATTACCTTAAAAAATAATTTTATAATTAAATAAAAAAAAAAAAAATATAAAAAAAATATCCAAAA : 211
CAII1 : TTTATTAATACATATTATATATTTTATTACCTTAAAAAATAATTTTATAATTAAATAAAAAAAAAAAAAATATAAAAAAAATATCCAAAA : 212
CAII2 : TTTATTAATACATATTATATATTTTATTACCTTAAAAAATAATTTTATAATTAAATAAAAAAAAAAAAAATATAAAAAAAATATCCAAAA : 213
EA : TTTATTAATACATATTATATATTTTATTACCTTAAAAAATAATTTTATAATTAAATAAAAAAAAAAAAAATATAAAAAAAATATCCAAAA : 212
Koren : TTTATTAATACATATTATATATTTTATTACCTTAAAAAATAATTTTATAATTAAATAAAAAAAAAAAAAATATAAAAAAAATATCCAAAA : -
Gene1990 : TTTATTAATACATATTATATATTTTATTACCTTAAAAAATAATTTTATAATTAAATAAAAAAAAAAAAAATATAAAAAAAATATCCAAAA : 212
Chao : TTTATTAATACATATTATATATTTTATTACCTTAAAAAATAATTTTATAATTAAATAAAAAAAAAAAAAATATAAAAAAAATATCCAAAA : -
      tttatgtaatacatgtattatatatttttggttaccgtgtaaaagtaattttatagttttaaaggagataaaaaaaaaa ttggat aaaaaaa tatccaaaat

      *      220      *      240      *      260      *      280      *      300      *      320
EU1 : CTATTTCATTATTATTAACTTTTTCATACTTTTAAACAACAATTTTCAAAATAAAAAATAACCTTCCTTTCTCTCTCTTAATTATCTATTTC : 320
EU2 : CTATTTCATTATTATTAACTTTTTCATACTTTTAAACAACAATTTTCAAAATAAAAAATAACCTTCCTTTCTCTCTCTTAATTATCTATTTC : 319
CAI1 : CTATTTCATTATTATTAACTTTTTCATACTTTTAAACAACAATTTTCAAAATAAAAAATAACCTTCCTTTCTCTCTCTTAATTATCTATTTC : 319
CAI2 : CTATTTCATTATTATTAACTTTTTCATACTTTTAAACAACAATTTTCAAAATAAAAAATAACCTTCCTTTCTCTCTCTTAATTATCTATTTC : 318
CAII1 : CTATTTCATTATTATTAACTTTTTCATACTTTTAAACAACAATTTTCAAAATAAAAAATAACCTTCCTTTCTCTCTCTTAATTATCTATTTC : 319
CAII2 : CTATTTCATTATTATTAACTTTTTCATACTTTTAAACAACAATTTTCAAAATAAAAAATAACCTTCCTTTCTCTCTCTTAATTATCTATTTC : 320
EA : CTATTTCATTATTATTAACTTTTTCATACTTTTAAACAACAATTTTCAAAATAAAAAATAACCTTCCTTTCTCTCTCTTAATTATCTATTTC : 319
Koren : CTATTTCATTATTATTAACTTTTTCATACTTTTAAACAACAATTTTCAAAATAAAAAATAACCTTCCTTTCTCTCTCTTAATTATCTATTTC : 36
Gene1990 : CTATTTCATTATTATTAACTTTTTCATACTTTTAAACAACAATTTTCAAAATAAAAAATAACCTTCCTTTCTCTCTCTTAATTATCTATTTC : 319
Chao : CTATTTCATTATTATTAACTTTTTCATACTTTTAAACAACAATTTTCAAAATAAAAAATAACCTTCCTTTCTCTCTCTTAATTATCTATTTC : 35
      ctgatttcatgggtatttggtaacttttccaggagcttgaacaacaatgtttccaataaaaaataaaccttcttcttctctctagATTAGTGCTATTGTGCG

      *      340      *      360      *      380      *      400      *      420
EU1 : TTTCTCTTTATTTTGTGAACCACTTAAATCATCAACAACCACTCTCTCTCAATAATCATCTATTCTTACAACACCTCACCACCTCTCTCAAA : 427
EU2 : TTTCTCTTTATTTTGTGAACCACTTAAATCATCAACAACCACTCTCTCTCAATAATCATCTATTCTTACAACACCTCACCACCTCTCTCAAA : 426
CAI1 : TTTCTCTTTATTTTGTGAACCACTTAAATCATCAACAACCACTCTCTCTCAATAATCATCTATTCTTACAACACCTCACCACCTCTCTCAAA : 426
CAI2 : TTTCTCTTTATTTTGTGAACCACTTAAATCATCAACAACCACTCTCTCTCAATAATCATCTATTCTTACAACACCTCACCACCTCTCTCAAA : 425
CAII1 : TTTCTCTTTATTTTGTGAACCACTTAAATCATCAACAACCACTCTCTCTCAATAATCATCTATTCTTACAACACCTCACCACCTCTCTCAAA : 426
CAII2 : TTTCTCTTTATTTTGTGAACCACTTAAATCATCAACAACCACTCTCTCTCAATAATCATCTATTCTTACAACACCTCACCACCTCTCTCAAA : 427
EA : TTTCTCTTTATTTTGTGAACCACTTAAATCATCAACAACCACTCTCTCTCAATAATCATCTATTCTTACAACACCTCACCACCTCTCTCAAA : 426
Koren : TTTCTCTTTATTTTGTGAACCACTTAAATCATCAACAACCACTCTCTCTCAATAATCATCTATTCTTACAACACCTCACCACCTCTCTCAAA : 143
Gene1990 : TTTCTCTTTATTTTGTGAACCACTTAAATCATCAACAACCACTCTCTCTCAATAATCATCTATTCTTACAACACCTCACCACCTCTCTCAAA : 426
Chao : TTTCTCTTTATTTTGTGAACCACTTAAATCATCAACAACCACTCTCTCTCAATAATCATCTATTCTTACAACACCTCACCACCTCTCTCAAA : 142
      GTGGTGTCTGGTTAGTTGTTGGT AACCAGGGGAGAGCATCAGACAAACAGCGGCTCTTCAATAATGCAGTCATTCTGTGTACAACCTGCACAGCTGGCTGCAAA

      *      440      *      460      *      480      *      500      *      520
EU1 : AATATTAACACTTTAAATTCTCATTAAATCTATTCAAACTAATAATTCTAAACTTAATTTAAACAAACATTATTATTCACACCTATTAA : 534
EU2 : AATATTAACACTTTAAATTCTCATTAAATCTATTCAAACTAATAATTCTAAACTTAATTTAAACAAACATTATTATTCACACCTATTAA : 533
CAI1 : AATATTAACACTTTAAATTCTCATTAAATCTATTCAAACTAATAATTCTAAACTTAATTTAAACAAACATTATTATTCACACCTATTAA : 533
CAI2 : AATATTAACACTTTAAATTCTCATTAAATCTATTCAAACTAATAATTCTAAACTTAATTTAAACAAACATTATTATTCACACCTATTAA : 532
CAII1 : AATATTAACACTTTAAATTCTCATTAAATCTATTCAAACTAATAATTCTAAACTTAATTTAAACAAACATTATTATTCACACCTATTAA : 533
CAII2 : AATATTAACACTTTAAATTCTCATTAAATCTATTCAAACTAATAATTCTAAACTTAATTTAAACAAACATTATTATTCACACCTATTAA : 534
EA : AATATTAACACTTTAAATTCTCATTAAATCTATTCAAACTAATAATTCTAAACTTAATTTAAACAAACATTATTATTCACACCTATTAA : 533
Koren : AATATTAACACTTTAAATTCTCATTAAATCTATTCAAACTAATAATTCTAAACTTAATTTAAACAAACATTATTATTCACACCTATTAA : 159
Gene1990 : AATATTAACACTTTAAATTCTCATTAAATCTATTCAAACTAATAATTCTAAACTTAATTTAAACAAACATTATTATTCACACCTATTAA : 533
Chao : AATATTAACACTTTAAATTCTCATTAAATCTATTCAAACTAATAATTCTAAACTTAATTTAAACAAACATTATTATTCACACCTATTAA : 159
      AATGATTACGACTTTgtaagattctcattaaatctatttcagaactagtagattctaaactgtaatttagagcaaacattattagtcacacctattaaagtttaaa

      *      540      *      560      *      580      *      600      *      620      *      640
EU1 : AAAAACTTTAACTCACCCTAAACCTCTATCAATCAATATTATTATATATCTATCAATATAAAATATTATATAAATAACAATTTTAAATACAA : 641
EU2 : AAAAACTTTAACTCACCCTAAACCTCTATCAATCAATATTATTATATATCTATCAATATAAAATATTATATAAATAACAATTTTAAATACAA : 640
CAI1 : AAAAACTTTAACTCACCCTAAACCTCTATCAATCAATATTATTATATATCTATCAATATAAAATATTATATAAATAACAATTTTAAATACAA : 640
CAI2 : AAAAACTTTAACTCACCCTAAACCTCTATCAATCAATATTATTATATATCTATCAATATAAAATATTATATAAATAACAATTTTAAATACAA : 639
CAII1 : AAAAACTTTAACTCACCCTAAACCTCTATCAATCAATATTATTATATATCTATCAATATAAAATATTATATAAATAACAATTTTAAATACAA : 640
CAII2 : AAAAACTTTAACTCACCCTAAACCTCTATCAATCAATATTATTATATATCTATCAATATAAAATATTATATAAATAACAATTTTAAATACAA : 641
EA : AAAAACTTTAACTCACCCTAAACCTCTATCAATCAATATTATTATATATCTATCAATATAAAATATTATATAAATAACAATTTTAAATACAA : 640
Koren : AAAAACTTTAACTCACCCTAAACCTCTATCAATCAATATTATTATATATCTATCAATATAAAATATTATATAAATAACAATTTTAAATACAA : -
Gene1990 : AAAAACTTTAACTCACCCTAAACCTCTATCAATCAATATTATTATATATCTATCAATATAAAATATTATATAAATAACAATTTTAAATACAA : 640
Chao : AAAAACTTTAACTCACCCTAAACCTCTATCAATCAATATTATTATATATCTATCAATATAAAATATTATATAAATAACAATTTTAAATACAA : -
      agtgaatgtgcttaggactcacctgtaaacctctatcagatgcaattatttattgagttatctgagtgttaaatgtattataaataagcaattttaaatacaat

      *      660      *      680      *      700      *      720      *      740
EU1 : ACTAAATACAATTTACCTTTTATATCAAAACAACCTTCTCTCTAACAAGCAACAATAAAAAATCTTCCCTCTCTTTTCTCAATTTCTACTACAF : 748
EU2 : ACTAAATACAATTTACCTTTTATATCAAAACAACCTTCTCTCTAACAAGCAACAATAAAAAATCTTCCCTCTCTTTTCTCAATTTCTACTACAF : 747
CAI1 : ACTAAATACAATTTGCGTTTATATCAAAACAACCTTCTCTCTAACAAGCAACAATAAAAAATCTTCCCTCTCTTTTCTCAATTTCTACTACAF : 747
CAI2 : ACTAAATACAATTTGCGTTTATATCAAAACAACCTTCTCTCTAACAAGCAACAATAAAAAATCTTCCCTCTCTTTTCTCAATTTCTACTACAF : 746
CAII1 : ACTAAATACAATTTGCGTTTATATCAAAACAACCTTCTCTCTAACAAGCAACAATAAAAAATCTTCCCTCTCTTTTCTCAATTTCTACTACAF : 747
CAII2 : ACTAAATACAATTTGCGTTTATATCAAAACAACCTTCTCTCTAACAAGCAACAATAAAAAATCTTCCCTCTCTTTTCTCAATTTCTACTACAF : 748
EA : ACTAAATACAATTTGCGTTTATATCAAAACAACCTTCTCTCTAACAAGCAACAATAAAAAATCTTCCCTCTCTTTTCTCAATTTCTACTACAF : 747
Koren : ACTAAATACAATTTGCGTTTATATCAAAACAACCTTCTCTCTAACAAGCAACAATAAAAAATCTTCCCTCTCTTTTCTCAATTTCTACTACAF : 236
Gene1990 : ACTAAATACAATTTGCGTTTATATCAAAACAACCTTCTCTCTAACAAGCAACAATAAAAAATCTTCCCTCTCTTTTCTCAATTTCTACTACAF : 747
Chao : ACTAAATACAATTTGCGTTTATATCAAAACAACCTTCTCTCTAACAAGCAACAATAAAAAATCTTCCCTCTCTTTTCTCAATTTCTACTACAF : 235
      actaaatacaattt ccttttgbttatcaaggAGGACAGCTGTGTGCTGAGGACCGAGACGCTGAGTAAAACTTCCTCTGTCTTTCTGCAATTCTGACTACAT
```


[illegible]

Alignment A. 3 (cont.)

```

00          *          1520      *          1540      *          1560      *          1580      *          1600
EU1 : AAAATGACATCACTTCTCATCCAACTTAA AAAACAAACAACTCTCTATCTCTACCAAACTTACACATCCTCTCAAATCTTATTTACCCCTCTAA : 1599
EU2 : AAAATGACATCACTTCTCATCCAACTTAA AAAACAAACAACTCTCTATCTCTACCAAACTTACACATCCTCTCAAATCTTATTTACCCCTCTAA : 1598
CAI1 : AAAATGACATCACTTCTCATCCAACTTAA AAAACAAACAACTCTCTATCTCTACCAAACTTACACATCCTCTCAAATCTTATTTACCCCTCTAA : 1602
CAI2 : AAAATGACATCACTTCTCATCCAACTTAA AAAACAAACAACTCTCTATCTCTACCAAACTTACACATCCTCTCAAATCTTATTTACCCCTCTAA : 1601
CAI11 : AAAATGACATCACTTCTCATCCAACTTAA AAAACAAACAACTCTCTATCTCTACCAAACTTACACATCCTCTCAAATCTTATTTACCCCTCTAA : 1261
CAI12 : AAAATGACATCACTTCTCATCCAACTTAA AAAACAAACAACTCTCTATCTCTACCAAACTTACACATCCTCTCAAATCTTATTTACCCCTCTAA : 1603
EA : AAAATGACATCACTTCTCATCCAACTTAA AAAACAAACAACTCTCTATCTCTACCAAACTTACACATCCTCTCAAATCTTATTTACCCCTCTAA : 1261
Koren : AAAATGACATCACTTCTCATCCAACTTAA AAAACAAACAACTCTCTATCTCTACCAAACTTACACATCCTCTCAAATCTTATTTACCCCTCTAA : 438
Gene1990 : AAAATGACATCACTTCTCATCCAACTTAA AAAACAAACAACTCTCTATCTCTACCAAACTTACACATCCTCTCAAATCTTATTTACCCCTCTAA : 1594
Chao : AAAATGACATCACTTCTCATCCAACTTAA AAAACAAACAACTCTCTATCTCTACCAAACTTACACATCCTCTCAAATCTTATTTACCCCTCTAA : 437
      AAAATGGGCATCAGTGTGCTCATCCAGgtgagaaagacaagcaagtgtctagtggctgctaccagaaggtgtgacagcatcctcaaa cttatttggag gctctaag

          *          1620      *          1640      *          1660      *          1680      *          1700
EU1 : TTCAATTTTACATTTTTCCTCACACATCTCTCATCTCAACCAAAACATATATAACACTCCTTCCGCTTCTTTTAACTTCTACTTACCATTTAG : 1706
EU2 : TTCAATTTTACATTTTTCCTCACACATCTCTCATCTCAACCAAAACATATATAACACTCCTTCCGCTTCTTTTAACTTCTACTTACCATTTAG : 1705
CAI1 : TTCAATTTTACATTTTTCCTCACACATCTCTCATCTCAACCAAAACATATATAACACTCCTTCCGCTTCTTTTAACTTCTACTTACCATTTAG : 1709
CAI2 : TTCAATTTTACATTTTTCCTCACACATCTCTCATCTCAACCAAAACATATATAACACTCCTTCCGCTTCTTTTAACTTCTACTTACCATTTAG : 1708
CAI11 : TTCAATTTTACATTTTTCCTCACACATCTCTCATCTCAACCAAAACATATATAACACTCCTTCCGCTTCTTTTAACTTCTACTTACCATTTAG : 1368
CAI12 : TTCAATTTTACATTTTTCCTCACACATCTCTCATCTCAACCAAAACATATATAACACTCCTTCCGCTTCTTTTAACTTCTACTTACCATTTAG : 1710
EA : TTCAATTTTACATTTTTCCTCACACATCTCTCATCTCAACCAAAACATATATAACACTCCTTCCGCTTCTTTTAACTTCTACTTACCATTTAG : 1368
Koren : -----CATCTCTCATCTCAACCAAAACATATATAACACTCCTTCCGCTTCTTTTAACTTCTACTTACCATTTAG : 519
Gene1990 : TTCAATTTTACATTTTTCCTCACACATCTCTCATCTCAACCAAAACATATATAACACTCCTTCCGCTTCTTTTAACTTCTACTTACCATTTAG : 1701
Chao : -----CATCTCTCATCTCAACCAAAACATATATAACACTCCTTCCGCTTCTTTTAACTTCTACTTACCATTTAG : 518
      tt attttacatttgtttctgcacagGcATGTCTCGATGGTCAACCAAAACATGGATGATAA GACTCCTTGCCTGCTGCTTTTGAGGAGCTTCTACTTGACCATGGGG

          *          1720      *          1740      *          1760      *          1780      *          1800      *          182
EU1 : AAACAAACCTCAAAATACTTTCCTCTCTCTTTCTTCAAAACACATCACAAAATCAGACCTACTTAAGTTCAAATTCAGAAATCCCTTAATTC : 1813
EU2 : AAACAAACCTCAAAATACTTTCCTCTCTCTTTCTTCAAAACACATCACAAAATCAGACCTACTTAAGTTCAAATTCAGAAATCCCTTAATTC : 1812
CAI1 : AAACAAACCTCAAAATACTTTCCTCTCTCTTTCTTCAAAACACATCACAAAATCAGACCTACTTAAGTTCAAATTCAGAAATCCCTTAATTC : 1816
CAI2 : AAACAAACCTCAAAATACTTTCCTCTCTCTTTCTTCAAAACACATCACAAAATCAGACCTACTTAAGTTCAAATTCAGAAATCCCTTAATTC : 1815
CAI11 : AAACAAACCTCAAAATACTTTCCTCTCTCTTTCTTCAAAACACATCACAAAATCAGACCTACTTAAGTTCAAATTCAGAAATCCCTTAATTC : 1475
CAI12 : AAACAAACCTCAAAATACTTTCCTCTCTCTTTCTTCAAAACACATCACAAAATCAGACCTACTTAAGTTCAAATTCAGAAATCCCTTAATTC : 1817
EA : AAACAAACCTCAAAATACTTTCCTCTCTCTTTCTTCAAAACACATCACAAAATCAGACCTACTTAAGTTCAAATTCAGAAATCCCTTAATTC : 1475
Koren : AAACAAACCTCAAAATACTTTCCTCTCTCTTTCTTCAAAACACATCACAAAATCAGACCTACTTAAGTTCAAATTCAGAAATCCCTTAATTC : 626
Gene1990 : AAACAAACCTCAAAATACTTTCCTCTCTCTTTCTTCAAAACACATCACAAAATCAGACCTACTTAAGTTCAAATTCAGAAATCCCTTAATTC : 1808
Chao : AAACAAACCTCAAAATACTTTCCTCTCTCTTTCTTCAAAACACATCACAAAATCAGACCTACTTAAGTTCAAATTCAGAAATCCCTTAATTC : 625
      GAGAACAACCTCAGAGAGCTTTGCTGTGCTGGCTTGCCTTcAAGAAGGACATGCACAAAGTCGagACCTActTGAGGGTTCGAAATTGCAGGAGATCCCTGGATTTC

0          *          1840      *          1860      *          1880      *          1900      *          1920
EU1 : CAACCTCACCCCTTAATTCACCCGGGTATTTTATTCAAATCCTTAACACATTTTCTTTCTCAAAATCTAAACCAATTAC----- : 1901
EU2 : CAACCTCACCCCTTAATTCACCCGGGTATTTTATTCAAATCCTTAACACATTTTCTTTCTCAAAATCTAAACCAATTAC----- : 1900
CAI1 : CAACCTCACCCCTTAATTCACCCGGGTATTTTATTCAAATCCTTAACACATTTTCTTTCTCAAAATATAAACCAATTAC----- : 1904
CAI2 : CAACCTCACCCCTTAATTCACCCGGGTATTTTATTCAAATCCTTAACACATTTTCTTTCTCAAAATATAAACCAATTAC----- : 1903
CAI11 : CAACCTCACCCCTTAATTCACCCGGGTATTTTATTCAAATCCTTAACACATTTTCTTTCTCAAAATATAAACCAATTAC----- : 1563
CAI12 : CAACCTCACCCCTTAATTCACCCGGGTATTTTATTCAAATCCTTAACACATTTTCTTTCTCAAAATATAAACCAATTAC----- : 1905
EA : CAACCTCACCCCTTAATTCACCCGGGTATTTTATTCAAATCCTTAACACATTTTCTTTCTCAAAATATAAACCAATTAC----- : 1563
Koren : CAACCTCACCCCTTAATTCACCCGGGTATTTTATTCAAATCCTTAACACATTTTCTTTCTCAAAATATAAACCAATTAC----- : 714
Gene1990 : CAACCTCACCCCTTAATTCACCCGGGTATTTTATTCAAATCCTTAACACATTTTCTTTCTCAAAATCTAAACCAATTAC----- : 1896
Chao : CAACCTCACCCCTTAATTCACCCGGGTATTTTATTCAAATCCTTAACACATTTTCTTTCTCAAAATCTAAACCAATTAC----- : 709
      CAACCTCACCCCTTAATTCACCCGGGTATTTTATTCAAATCCTTAACACATTTTCTTTCTCAAAATCTAAACCAATTAC-----

          *          1940      *          1960      *          1980      *          2000      *          2020      *
EU1 : -----AATCTTCTCTTATATGTCAAAATTCACACAGCATCCTAGTCTTGTTTTCTTTTTCCCTCCCATATTTAAACCTTACCTA----- : 1994
EU2 : -----AATCTTCTCTTATATGTCAAAATTCACACAGCATCCTAGTCTTGTTTTCTTTTTCCCTCCCATATTTAAACCTTACCTA----- : 1993
CAI1 : -----AGTCTTCTCTTATATGTCAAAATTCACACAGCATCCTAGTCTTGTTTTCTTTTTCCCTCCCATATTTAAACCTTACCTA----- : 1997
CAI2 : -----AGTCTTCTCTTATATGTCAAAATTCACACAGCATCCTAGTCTTGTTTTCTTTTTCCCTCCCATATTTAAACCTTACCTA----- : 1996
CAI11 : -----AATCTTCTCTTATATGTCAAAATTCACACAGCATCCTAGTCTTGTTTTCTTTTTCCCTCCCATATTTAAACCTTACCTA----- : 1656
CAI12 : -----AATCTTCTCTTATATGTCAAAATTCACACAGCATCCTAGTCTTGTTTTCTTTTTCCCTCCCATATTTAAACCTTACCTA----- : 1998
EA : -----AATCTTCTCTTATATGTCAAAATTCACACAGCATCCTAGTCTTGTTTTCTTTTTCCCTCCCATATTTAAACCTTACCTA----- : 1656
Koren : -----AGTCTTCTCTTATATGTCAAAATTCACACAGCATCCTAGTCTTGTTTTCTTTTTCCCTCCCATATTTAAACCTTACCTA----- : 806
Gene1990 : -----AATCTTCTCTTATATGTCAAAATTCACACAGCATCCTAGTCTTGTTTTCTTTTTCCCTCCCATATTTAAACCTTACCTA----- : 1989
Chao : TAAATATGATCTTCTCTTATATGTCAAAATTCACACAGCATCCTAGTCTTGTTTTCTTTTTCCCTCCCATATTTAAACCTTATCTACCACTATT : 815
      a TCTGGTCTTATATgTGcAGGAAaTGTCAACCAgGcATGcCTAGTgCTGTgTTTTCTTTTTcCCCTCCcATATTTTAAAC TTAcCTA TTGTA

          *          2040      *          2060      *          2080      *          2100      *          2120
EU1 : TTATTCTCTCATTTTAAAGTCTCATAAATTTAAACCTTTCCCTTTAAACCTTAAGGATATATCTTAACATTTCACATTTT : 2080
EU2 : TTATTCTCTCATTTTAAAGTCTCATAAATTTAAACCTTTCCCTTTAAACCTTAAGGATATATCTTAACATTTCACATTTT : 2079
CAI1 : TTATTCTCTCATTTTAAAGTCTCATAAATTTAAACCTTTCCCTTTAAACCTTAAGGATATATCTTAACATTTCACATTTT : 2083
CAI2 : TTATTCTCTCATTTTAAAGTCTCATAAATTTAAACCTTTCCCTTTAAACCTTAAGGATATATCTTAACATTTCACATTTT : 2082
CAI11 : TTATTCTCTCATTTTAAAGTCTCATAAATTTAAACCTTTCCCTTTAAACCTTAAGGATATATCTTAACATTTCACATTTT : 1742
CAI12 : TTATTCTCTCATTTTAAAGTCTCATAAATTTAAACCTTTCCCTTTAAACCTTAAGGATATATCTTAACATTTCACATTTT : 2084
EA : TTATTCTCTCATTTTAAAGTCTCATAAATTTAAACCTTTCCCTTTAAACCTTAAGGATATATCTTAACATTTCACATTTT : 1742
Koren : TTATTCTCTCATTTTAAAGTCTCATAAATTTAAACCTTTCCCTTTAAACCTTAAGGATATATCTTAACATTTCACATTTT : 894
Gene1990 : TTATTCTCTCATTTTAAAGTCTCATAAATTTAAACCTTTCCCTTTAAACCTTAAGGATATATCTTAACATTTCACATTTT : 2075
Chao : TTATTCTCTCATTTTAAAGTCTCATAAATTTAAACCTTTCCCTTTAAACCTTAAGGATATATCTTAACATTTCACATTTT : 897
      TTTATTCTCTCATTTTGGGagTG CTCaTAAATTT AAAaCcGTTCCCTTTAAACCTAAGGATGGATCTGgaacATTTCAAGTGGTG

```

Alignment A. 4 Alignment of *GH II* gene fragments. First fragment from first exon to second exon obtained with primer pair GH-a + GH-k: EU - sample from Europe (River Rhine), clones 1 and 2; CA I - clones 1 and 2, CA II - clones 1 and 2, two samples from Central Asia (Uzbekistan); EA - sample from East Asia (China), clones 1 and 2. Second fragment from second exon to the beginning of the third intron obtained with primer pair GH-g + GH-h: 2EU, 2CA I, 2CA II, 2EA. Third fragment from the beginning of the third intron to 3' non-coding end obtained with primer pair GH-e + GH-b: 3EU, 3CA I, 3CA II 3EA. Koren, 2Koren, 3Koren - cDNA from Koren et al. (1989) (*GH I*); Chao, 2Chao, 3Chao - cDNA from Chao et al. (1989) (*GH II*). See also Figure II.1 in Chapter II of this study.

			*	20		*	40		*	60		*	80		*	100			
Koren	:	T	A	C	A	A	A	T	C	T	A	A	-----	-----	-----	-----	-----	18	
Chao	:	T	A	C	A	A	A	T	C	T	A	A	-----	-----	-----	-----	-----	18	
EU1	:	T	A	C	A	A	A	T	C	T	A	A	T	A	A	A	C	T	106
EU2	:	T	A	C	A	A	A	T	C	T	A	A	T	A	A	A	C	T	106
CAI1	:	T	A	C	A	A	A	T	C	T	A	A	T	A	A	A	C	T	92
CAI2	:	T	A	C	A	A	A	T	C	T	A	A	T	A	A	A	C	T	92
CAII1	:	T	A	C	A	A	A	T	C	T	A	A	T	A	A	A	C	T	92
CAII2	:	T	A	C	A	A	A	T	C	T	A	A	T	A	A	A	C	T	92
EA1	:	T	A	C	A	A	A	T	C	T	A	A	T	A	A	A	C	T	107
EA2	:	T	A	C	A	A	A	T	C	T	A	A	T	A	A	A	C	T	107
2Koren	:	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-	
2Chao	:	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-	
2EU1	:	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-	
2EU2	:	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-	
2CAI1,2	:	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-	
2CAII1,2	:	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-	
2EA1	:	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-	
2EA2	:	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-	
3Koren	:	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-	
3Chao	:	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-	
3EU1	:	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-	
3EU2	:	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-	
3CAI1	:	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-	
3CAI2	:	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-	
3CAII1	:	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-	
3CAII2	:	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-	
3EA1	:	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-	
3EA2	:	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-	
	:	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-	
			*	120		*	140		*	160		*	180		*	200		*	
Koren	:	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-	
Chao	:	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-	
EU1	:	-----	T	A	T	A	T	A	T	A	T	A	T	A	T	A	T	176	
EU2	:	-----	T	A	T	A	T	A	T	A	T	A	T	A	T	A	T	180	
CAI1	:	-----	T	A	T	A	T	A	T	A	T	A	T	A	T	A	T	159	
CAI2	:	-----	T	A	T	A	T	A	T	A	T	A	T	A	T	A	T	163	
CAII1	:	-----	T	A	T	A	T	A	T	A	T	A	T	A	T	A	T	160	
CAII2	:	-----	T	A	T	A	T	A	T	A	T	A	T	A	T	A	T	162	
EA1	:	-----	A	C	A	C	A	C	A	T	A	T	A	T	A	T	A	210	
EA2	:	-----	A	C	A	C	A	C	A	T	A	T	A	T	A	T	A	208	
2Koren	:	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-	
2Chao	:	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-	
2EU1	:	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-	
2EU2	:	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-	
2CAI1,2	:	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-	
2CAII1,2	:	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-	
2EA1	:	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-	
2EA2	:	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-	
3Koren	:	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-	
3Chao	:	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-	
3EU1	:	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-	
3EU2	:	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-	
3CAI1	:	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-	
3CAI2	:	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-	
3CAII1	:	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-	
3CAII2	:	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-	
3EA1	:	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-	
3EA2	:	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-	
	:	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-	

Alignment A. 2 (continued)

	220	*	240	*	260	*	280	*	300	*	320		
Koren :	-----											:	
Chao :	-----											:	
EU1 :	TTAATTAA	TTAATTATTATT	TTTTTTTTT	TACA	AAAAAAAAA	TTTATT	AA	CAT	AAAAAAAA	CCCTTATCCA	:	271	
EU2 :	TTAATTAA	TTAATTATTATT	TTTTTTTTT	TACA	AAAAAAAAA	TTTATT	AA	CAT	AAAAAAAA	CCCTTATCCA	:	276	
CAI1 :	TTAATTAA	TGAATTATTATTATT	TTTTTTTTT	TACA	AAA	TTTTTTTTT	AAAA	TTTATT	AA	CAT	AAAAAAAA	CCCTTATCCA	: 254
CAI2 :	TTAATTAA	TGAATTATTATTATT	TTTTTTTTT	TACA	AAA	TTTTTTTTT	AAAA	TTTATT	AA	CAT	AAAAAAAA	CCCTTATCCA	: 258
CAII1 :	TTAATTAA	TGAATTATTATTATT	TTTTTTTTT	TACA	AAA	TTTTTTTTT	AAAA	TTTATT	AA	CAT	AAAAAAAA	CCCTTATCCA	: 255
CAII2 :	TTAATTAA	TGAATTATTATTATT	TTTTTTTTT	TACA	AAA	TTTTTTTTT	AAAA	TTTATT	AA	CAT	AAAAAAAA	CCCTTATCCA	: 257
EA1 :	TTAATTAA	TGAATTATTATTATT	TTTTTTTTT	TACA	AAAA	TTTTTTTTT	AAAA	TTTATT	AA	CAT	AAAAAAAA	CCCTTATCCA	: 307
EA2 :	TTAATTAA	TGAATTATTATTATT	TTTTTTTTT	TACA	AAAA	TTTTTTTTT	AAAA	TTTATT	AA	CAT	AAAAAAAA	CCCTTATCCA	: 307
2Koren :	-----											:	
2Chao :	-----											:	
2EU1 :	-----											:	
2EU2 :	-----											:	
2CAI1,2 :	-----											:	
2CAII1,2 :	-----											:	
2EA1 :	-----											:	
2EA2 :	-----											:	
3Koren :	-----											:	
3Chao :	-----											:	
3EU1 :	-----											:	
3EU2 :	-----											:	
3CAI1 :	-----											:	
3CAI2 :	-----											:	
3CAII1 :	-----											:	
3CAII2 :	-----											:	
3EA1 :	-----											:	
3EA2 :	-----											:	

[illegible][illegible]

Alignment A. 2 (cont.)

		540	*	560	*	580	*	600	*	620	*	640							
Koren	:	CTACCC	CAAAAAAT	ATTAAG	ACTTT								157						
Chao	:	CTACCC	CAAAAAAT	ATTAAG	ACTTT								158						
EU1	:	CTACCC	CAAAAAAT	ATTAAG	ACTTT	TAA							517						
EU2	:	CTACCC	CAAAAAAT	ATTAAG	ACTTT	TAA							522						
CAI1	:	CTACCC	CAAAAAAT	ATTAAG	ACTTT	TAA							499						
CAI2	:	CTACCC	CAAAAAAT	ATTAAG	ACTTT	TAA							503						
CAII1	:	CTACCC	CAAAAAAT	ATTAAG	ACTTT	TAA							500						
CAII2	:	CTACCC	CAAAAAAT	ATTAAG	ACTTT	TAA							502						
EA1	:	CTACCC	CAAAAAAT	ATTAAG	ACTTT	TAA							504						
EA2	:	CTACCC	CAAAAAAT	ATTAAG	ACTTT	TAA							504						
2Koren	:	CTACCT	CAAAAAAT	ATTAAG	ACTTT								77						
2Chao	:	CTACCC	CAAAAAAT	ATTAAG	ACTTT								77						
2EU1	:	CTACCC	CAAAAAAT	ATTAAG	ACTTT	TAA	ATTCTCAACAAAATCTATTCA	AACTA	TAA	ATTCTAAACCTCTCATTTACA	CAAAAT	TAA	TCACACCT	157					
2EU2	:	CTACCC	CAAAAAAT	ATTAAG	ACTTT	TAA	ATTCTCAACAAAATCTATTCA	AACTA	TAA	ATTCTAAACCTCTCATTTACA	CAAAAT	TAA	TCACACCT	122					
2CAI1,2	:	CTACCC	CAAAAAAT	ATTAAG	ACTTT	TAA	ATTCTCAACAAAATCTATTCA	AACTA	TAA	ATTCTAAACCTCTCATTTACA	CAAAAT	TAA	TCACACCT	157					
2CAII1,2	:	CTACCC	CAAAAAAT	ATTAAG	ACTTT	TAA	ATTCTCAACAAAATCTATTCA	AACTA	TAA	ATTCTAAACCTCTCATTTACA	CAAAAT	TAA	TCACACCT	157					
2EA1	:	CTACCC	CAAAAAAT	ATTAAG	ACTTT	TAA	ATTCTCAACAAAATCTATTCA	AACTA	TAA	ATTCTAAACCTCTCATTTACA	CAAAAT	TAA	TCACACCT	157					
2EA2	:	CTACCC	CAAAAAAT	ATTAAG	ACTTT	TAA	ATTCTCAACAAAATCTATTCA	AACTA	TAA	ATTCTAAACCTCTCATTTACA	CAAAAT	TAA	TCACACCT	157					
3Koren	:													-					
3Chao	:													-					
3EU1	:													-					
3EU2	:													-					
3CAI1	:													-					
3CAI2	:													-					
3CAII1	:													-					
3CAII2	:													-					
3EA1	:													-					
3EA2	:													-					
	:													-					
		*	660	*	680	*	700	*	720	*	740								
Koren	:													-					
Chao	:													-					
EU1	:													-					
EU2	:													-					
CAI1	:													-					
CAI2	:													-					
CAII1	:													-					
CAII2	:													-					
EA1	:													-					
EA2	:													-					
2Koren	:													-					
2Chao	:													-					
2EU1	:	TCAC	TTTAAAAAT	ATATAT	ACTTA	ACTTT	AA	AACTCATCTAA	CACCTCT	CA	AT	CATTATTTAATTACCTTA	TAA	TTAAAT	TATTATAAATA	264			
2EU2	:	TCAC	TTTAAAAAT	ATATAT	ACTTA	ACTTT	AA	AACTCATCTAA	CACCTCT	CA	AT	CATTATTTAATTACCTTA	TAA	TTAAAT	TTTATAAATA	229			
2CAI1,2	:	TCAC	TTTAAAAAT	ATATAT	ACTTA	ACTTT	AA	AACTCATCTAA	CACCTCT	CA	AT	CATTATTTAATTACCTTA	TAA	TTAAAT	TTTATAAATA	264			
2CAII1,2	:	TCAC	TTTAAAAAT	ATATAT	ACTTA	ACTTT	AA	AACTCATCTAA	CACCTCT	CA	AT	CATTATTTAATTACCTTA	TAA	TTAAAT	TTTATAAATA	264			
2EA1	:	TCAC	TTTAAAAAT	ATATAT	ACTTA	ACTTT	AA	AACTCATCTAA	CACCTCT	CA	AT	CATTATTTAATTACCTTA	TAA	TTAAAT	TTTATAAATA	264			
2EA2	:	TCAC	TTTAAAAAT	ATATAT	ACTTA	ACTTT	AA	AACTCATCTAA	CACCTCT	CA	AT	CATTATTTAATTACCTTA	TAA	TTAAAT	TTTATAAATA	264			
3Koren	:														-				
3Chao	:														-				
3EU1	:														-				
3EU2	:														-				
3CAI1	:														-				
3CAI2	:														-				
3CAII1	:														-				
3CAII2	:														-				
3EA1	:														-				
3EA2	:														-				
	:														-				
		*	760	*	780	*	800	*	820	*	840	*							
Koren	:														-				
Chao	:														-				
EU1	:														-				
EU2	:														-				
CAI1	:														-				
CAI2	:														-				
CAII1	:														-				
CAII2	:														-				
EA1	:														-				
EA2	:														-				
2Koren	:														-				
2Chao	:														-				
2EU1	:	AT	CATTTCAAAT	CAATACTAAATACAATT	CCITTT	TTATCAAA	AA	ATA	CCT	TT	CCCA	AGC	CA	ACA	CT	A	TAAAAATCTTCGCTCT	TCCTTT	136
2EU2	:	AT	CATTTCAAAT	CAATACTAAATACAATT	CCITTT	TTATCAAA	AA	ATA	CCT	TT	CCCA	AGC	CA	ACA	CT	A	TAAAAATCTTCGCTCT	TCCTTT	136
2CAI1,2	:	AT	CATTTCAAAT	CAATACTAAATACAATT	CCITTT	TTATCAAA	AA	ATA	CCT	TT	CCCA	AGC	CA	ACA	CT	A	TAAAAATCTTCGCTCT	TCCTTT	371
2CAII1,2	:	AT	CATTTCAAAT	CAATACTAAATACAATT	CCITTT	TTATCAAA	AA	ATA	CCT	TT	CCCA	AGC	CA	ACA	CT	A	TAAAAATCTTCGCTCT	TCCTTT	371
2EA1	:	AT	CATTTCAAAT	CAATACTAAATACAATT	CCITTT	TTATCAAA	AA	ATA	CCT	TT	CCCA	AGC	CA	ACA	CT	A	TAAAAATCTTCGCTCT	TCCTTT	371
2EA2	:	AT	CATTTCAAAT	CAATACTAAATACAATT	CCITTT	TTATCAAA	AA	ATA	CCT	TT	CCCA	AGC	CA	ACA	CT	A	TAAAAATCTTCGCTCT	TCCTTT	371
3Koren	:																	-	
3Chao	:																	-	
3EU1	:																	-	
3EU2	:																	-	
3CAI1	:																	-	
3CAI2	:																	-	
3CAII1	:																	-	
3CAII2	:																	-	
3EA1	:																	-	
3EA2	:																	-	
	:																	-	

Alignment A. 2(cont.)

	860	*	880	*	900	*	920	*	940	*	960
Koren	:										
Chao	:										
EU1	:										
EU2	:										
CAI1	:										
CAI2	:										
CAII1	:										
CAII2	:										
EA1	:										
EA2	:										
2Koren	:	CT	CAATTCT	ACTACATT	AATG	CCCTGCT	AAAAAT	AAACACA	AAAATCTCT		
2Chao	:	CT	CAATTCT	ACTACATT	AATG	CCCACT	AAAAAT	AAACACA	AAAATCTCT		
2EU1	:	CT	CAATTCT	ACTACATT	AATG	CCCACT	AAAAAT	AAACACA	AAAATCTCT	TAAATAAACTATATCAT	TAATAAAATGAAATCCCAACATCTTTTTCCTCA
2EU2	:	CT	CAATTCT	ACTACATT	AATG	CCCACT	AAAAAT	AAACACA	AAAATCTCT	TAAATAAACTATATCAT	TAATAAAATGAAATCCCAACATCTTTTTCCTCA
2CAI1,2	:	CT	CAATTCT	ACTACATT	AATG	CCCACT	AAAAAT	AAACACA	AAAATCTCT	TAAATAAACTATATCAT	TAATAAAATGAAATCCCAACATCTTTTTCCTCA
2CAII1,2	:	CT	CAATTCT	ACTACATT	AATG	CCCACT	AAAAAT	AAACACA	AAAATCTCT	TAAATAAACTATATCAT	TAATAAAATGAAATCCCAACATCTTTTTCCTCA
2EA1	:	CT	CAATTCT	ACTACATT	AATG	CCCACT	AAAAAT	AAACACA	AAAATCTCT	TAAATAAACTATATCAT	TAATAAAATGAAATCCCAACATCTTTTTCCTCA
2EA2	:	CT	CAATTCT	ACTACATT	AATG	CCCACT	AAAAAT	AAACACA	AAAATCTCT	TAAATAAACTATATCAT	TAATAAAATGAAATCCCAACATCTTTTTCCTCA
3Koren	:										
3Chao	:										
3EU1	:									AAAAATGAAATCCCAACATCTTTTTCCTCA	
3EU2	:									AAAAATGAAATCCCAACATCTTTTTCCTCA	
3CAI1	:									AAAAATGAAATCCCAACATCTTTTTCCTCA	
3CAI2	:									AAAAATGAAATCCCAACATCTTTTTCCTCA	
3CAII1	:									AAAAATGAAATCCCAACATCTTTTTCCTCA	
3CAII2	:									AAAAATGAAATCCCAACATCTTTTTCCTCA	
3EA1	:									AAAAATTTAATCCCAACATCTTTTTCCTCA	
3EA2	:									AAAAATTTAATCCCAACATCTTTTTCCTCA	

		*	980	*	1000	*	1020	*	1040	*	1060	*
Koren	:	----		----		----		----		----		----
Chao	:	----		----		----		----		----		----
EU1	:	----		----		----		----		----		----
EU2	:	----		----		----		----		----		----
CAI1	:	----		----		----		----		----		----
CAI2	:	----		----		----		----		----		----
CAI1,1	:	----		----		----		----		----		----
CAI1,2	:	----		----		----		----		----		----
EA1	:	----		----		----		----		----		----
EA2	:	----		----		----		----		----		----
2Koren	:	----		----		----		----		----		----
2Chao	:	----		----		----		----		----		----
2EU1	:	----		----		----		----		----		----
2EU2	:	----		----		----		----		----		----
2CAI1,2	:	----		----		----		----		----		----
2CAI1,1,2	:	----		----		----		----		----		----
2EA1	:	----		----		----		----		----		----
2EA2	:	----		----		----		----		----		----
3Koren	:	----		----		----		----		----		----
3Chao	:	----		----		----		----		----		----
3EU1	:	----		----		----		----		----		----
3EU2	:	----		----		----		----		----		----
3CAI1	:	----		----		----		----		----		----
3CAI2	:	----		----		----		----		----		----
3CAI1,1	:	----		----		----		----		----		----
3CAI1,2	:	----		----		----		----		----		----
3EA1	:	----		----		----		----		----		----
3EA2	:	----		----		----		----		----		----

	1080	*	1100	*	1120	*	1140	*	1160	*	1					
Koren	:															
Chao	:															
EU1	:															
EU2	:															
CAI1	:															
CAI2	:															
CAII1	:															
CAII2	:															
EA1	:															
EA2	:															
2Koren	:															
2Chao	:															
2EU1	:															
2EU2	:															
2CAI1,2	:															
2CAII1,2	:															
2EA1	:															
2EA2	:															
3Koren	:															
3Chao	:															
3EU1	:	AATCA	TAATAACAAATAGAATAA	TAAAAAATTATA	AAA											
3EU2	:	AATCA	TAATAACAAATAGAATAA	TAAAAAATTATA	AAA											
3CAI1	:	AATCA	TAATAACAAATAAAATAA	TAAAAAATAATA	AAA	CT	CATT	AAATACATCTT	TACTATA	TAA	CAAAAAA	TAA	AAATATAA	TAA	TAAAAA	T
3CAI2	:	AATCA	TAATAACAAATAAAATAA	TAAAAAATAATA	AAA	CT	CATT	AAATACATCTT	TACTATA	TAA	CAAAAAA	TAA	AAATATAA	TAA	TAAAAA	T
3CAII1	:	AATCA	TAATAACAAATAAAATAA	TAAAAAATAATA	AAA	CT	CATT	AAATACATCTT	TACTATA	TAA	CAAAAAA	TAA	AAATATAA	TAA	TAAAAA	T
3CAII2	:	AATCA	TAATAACAAATAAAATAA	TAAAAAATAATA	AAA	CT	CATT	AAATACATCTT	TACTATA	TAA	CAAAAAA	TAA	AAATATAA	TAA	TAAAAA	T
3EA1	:	AATCA	TAATAACAAATAAAATAA	TAAAAAATAATA	AAA	CT	CATT	AAATACATCTT	TACTATA	TAA	CAAAAAA	TAA	AAATATAA	TAA	TAAAAA	T
3EA2	:	AATCA	TAATAACAAATAAAATAA	TAAAAAATAATA	AAA	CT	CATT	AAATACATCTT	TACTATA	TAA	CAAAAAA	TAA	AAATATAA	TAA	TAAAAA	T

Alignment A. 2(cont.)

[illegible]

		*	1300	*	1320	*	1340	*	1360	*	1380	*	
Koren	:												:
Chao	:												:
EU1	:												:
EU2	:												:
CAI1	:												:
CAI2	:												:
CAI11	:												:
CAI12	:												:
EA1	:												:
EA2	:												:
2Koren	:												:
2Chao	:												:
2EU1	:												:
2EU2	:												:
2CAI1,2	:												:
2CAI11,2	:												:
2EA1	:												:
2EA2	:												:
3Koren	:												:
3Chao	:												:
3EU1	:												:
3EU2	:												:
3CAI1	:	--	CTCCTTTTAAATTTTA	AAgATTA	AAAAAAAAAAAAAA	TTCAAT	TAATATATACTATATTAT	AAACAAAT	TATCAAAACAAAT	AAAAAAA	TACCAACATTA		: 439
3CAI2	:	--	CTCCTTTTAAATTTTA	AAAAATTA	AAAAAAAAAAAAAA	TTCAAT	TAATATATACTATATTAT	AAACAAAT	TATCAAAACAAAT	AAAAAAA	TACCAACATTA		: 439
3CAI11	:	--	CTCCTTTTAAATTTTA	AAAAATTA	AAAAAAAAAAAAAA	TTCAAT	TAATATATACTATATTAT	AAACAAAT	TATCAAAACAAAT	AAAAAAA	TACCAACATTA		: 440
3CAI12	:	--	CTCCTTTTAAATTTTA	AAAAATTA	AAAAAAAAAAAAAA	TTCAAT	TAATATATACTATATTAT	AAACAAAT	TATCAAAACAAAT	AAAAAAA	TACCAACATTA		: 440
3EA1	:	TA	CTCCTTTTAAATTTTA	AAAAATTA	AAAAAAAAAAAAAA	TTCAAT	TAATATATACTATATTAT	AAACAAAT	TATCAAAACAAAT	AAAAAAA	TACCAACATTA		: 458
3EA2	:	TA	CTCCTTTTAAATTTTA	AAAAATTA	AAAAAAAAAAAAAA	TTCAAT	TAATATATACTATATTAT	AAACAAAT	TATCAAAACAAAT	AAAAAAA	TACCAACATTA		: 456

[illegible]

Alignment A. 2(cont.)

	00	*	1520	*	1540	*	1560	*	1580	*	1600	
Koren	:	----	----	----	----	----	----	----	----	----	----	:
Chao	:	----	----	----	----	----	----	----	----	----	----	:
EU1	:	----	----	----	----	----	----	----	----	----	----	:
EU2	:	----	----	----	----	----	----	----	----	----	----	:
CAI1	:	----	----	----	----	----	----	----	----	----	----	:
CAI2	:	----	----	----	----	----	----	----	----	----	----	:
CAI1,1	:	----	----	----	----	----	----	----	----	----	----	:
CAI1,2	:	----	----	----	----	----	----	----	----	----	----	:
EA1	:	----	----	----	----	----	----	----	----	----	----	:
EA2	:	----	----	----	----	----	----	----	----	----	----	:
2Koren	:	----	----	----	----	----	----	----	----	----	----	:
2Chao	:	----	----	----	----	----	----	----	----	----	----	:
2EU1	:	----	----	----	----	----	----	----	----	----	----	:
2EU2	:	----	----	----	----	----	----	----	----	----	----	:
2CAI1,2	:	----	----	----	----	----	----	----	----	----	----	:
2CAI1,1,2	:	----	----	----	----	----	----	----	----	----	----	:
2EA1	:	----	----	----	----	----	----	----	----	----	----	:
2EA2	:	----	----	----	----	----	----	----	----	----	----	:
3Koren	:	----	----	----	----	----	----	----	----	----	----	:
3Chao	:	----	----	----	----	----	----	----	----	----	----	:
3EU1	:	AACCACA	TAA	ATCTAA	A	ATTAAATTCTCCAT	CATAACAT	TTTTTTTTTTTTTTTT	TTT	AA	CATAAACATCAAAAAA	CC
3EU2	:	AACCACA	TAA	ATCTAA	A	ATTAAATTCTCCAT	CATAACAT	TTTTTTTTTTTTTTTT	TTT	AA	CATAAACATCAAAAAA	CC
3CAI1	:	AACCACA	TAA	ATCTAA	A	ATTAAATTCTCCAT	CATAACAC	TTTTTTTTTTTTTTTT	TTT	AA	CATAAACATCAAAAAA	CC
3CAI2	:	AACCACA	TAA	ATCTAA	A	ATTAAATTCTCCAT	CATAACAC	TTTTTTTTTTTTTTTT	TTT	AA	CATAAACATCAAAAAA	CC
3CAI1,1	:	AACCACA	TAA	ATCTAA	A	ATTAAATTCTCCAT	CATAACAC	TTTTTTTTTTTTTTTT	TTT	AA	CATAAACATCAAAAAA	CC
3CAI1,2	:	AACCACA	TAA	ATCTAA	A	ATTAAATTCTCCAT	CATAACAC	TTTTTTTTTTTTTTTT	TTT	AA	CATAAACATCAAAAAA	CC
3EA1	:	AACCACA	TAA	ATCTAA	A	ATTAAATTCTCCAT	CATAACAC	TGTTTTTTTTTTTTTTTT	TTT	AA	CATAAACATCAAAAAA	CC
3EA2	:	AACCACA	TAA	ATCTAA	A	ATTAAATTCTCCAT	CATAACACTG	TTTTTTTTTTTTTTTT	TTT	AA	CATAAACATCAAAAAA	CC

[illegible]

	1720	*	1740	*	1760	*	1780	*	1800	*	182	
Koren	:	:	:	:	:	:	:	:	:	:	:	-
Chao	:	:	:	:	:	:	:	:	:	:	:	-
EU1	:	:	:	:	:	:	:	:	:	:	:	-
EU2	:	:	:	:	:	:	:	:	:	:	:	-
CAI1	:	:	:	:	:	:	:	:	:	:	:	-
CAI2	:	:	:	:	:	:	:	:	:	:	:	-
CAII1	:	:	:	:	:	:	:	:	:	:	:	-
CAII2	:	:	:	:	:	:	:	:	:	:	:	-
EA1	:	:	:	:	:	:	:	:	:	:	:	-
EA2	:	:	:	:	:	:	:	:	:	:	:	-
2Koren	:	:	:	:	:	:	:	:	:	:	:	-
2Chao	:	:	:	:	:	:	:	:	:	:	:	-
2EU1	:	:	:	:	:	:	:	:	:	:	:	-
2EU2	:	:	:	:	:	:	:	:	:	:	:	-
2CAI1,2	:	:	:	:	:	:	:	:	:	:	:	-
2CAII1,2	:	:	:	:	:	:	:	:	:	:	:	-
2EA1	:	:	:	:	:	:	:	:	:	:	:	-
2EA2	:	:	:	:	:	:	:	:	:	:	:	-
3Koren	C	A	A	C	T	C	T	C	A	A	A	: 162
3Chao	C	A	A	C	T	C	T	C	A	A	A	: 162
3EU1	C	A	A	C	T	C	T	C	A	A	A	: 596
3EU2	C	A	A	C	T	C	T	C	A	A	A	: 594
3CAI1	C	A	A	C	T	C	T	C	A	A	A	: 856
3CAI2	C	A	A	C	T	C	T	C	A	A	A	: 857
3CAII1	C	A	A	C	T	C	T	C	A	A	A	: 860
3CAII2	C	A	A	C	T	C	T	C	A	A	A	: 859
3EA1	C	A	A	C	T	C	T	C	A	A	A	: 884
3EA2	C	A	A	C	T	C	T	C	A	A	A	: 884

Alignment A. 2(cont.)

	0	*	1840	*	1860	*	1880	*	1900	*	1920			
Koren	:											:		
Chao	:											:		
EU1	:											:		
EU2	:											:		
CAI1	:											:		
CAI2	:											:		
CAII1	:											:		
CAII2	:											:		
EA1	:											:		
EA2	:											:		
2Koren	:											:		
2Chao	:											:		
2EU1	:											:		
2EU2	:											:		
2CAI1,2	:											:		
2CAII1,2	:											:		
2EA1	:											:		
2EA2	:											:		
3Koren	:											:		
3Chao	:											:		
3EU1	:	A	CTTCTA	CTACTTTATTTTTT	CACACAACCTT	TCTA	ATG	TACCA	AAATTT	ACACATCTTCACATCTTCAA	CTTATTTT	AGCATATAA	AT	703
3EU2	:	A	CTTCTA	CTACTTTATTTTTT	CACACAACCTT	TCTA	ATG	TACCA	AAATTT	ACACATCTTCACATCTTCAA	CTTATTTT	AGCATATAA	AT	700
3CAI1	:	A	CTTCTA	CTACTTTATTTTTT	CACACAACCTT	TCTA	ATG	TACCA	AAATTT	ACACATCTTCACATCTTCAA	CTTATTTT	AGCATATAA	AT	962
3CAI2	:	A	CTTCTA	CTACTTTATTTTTT	CACACAACCTT	TCTA	ATG	TACCA	AAATTT	ACACATCTTCACATCTTCAA	CTTATTTT	AGCATATAA	AT	963
3CAII1	:	A	CTTCTA	CTACTTTATTTTTT	CACACAACCTT	TCTA	ATG	TACCA	AAATTT	ACACATCTTCACATCTTCAA	CTTATTTT	AGCATATAA	AT	966
3CAII2	:	A	CTTCTA	CTACTTTATTTTTT	CACACAACCTT	TCTA	ATG	TACCA	AAATTT	ACACATCTTCACATCTTCAA	CTTATTTT	AGCATATAA	AT	965
3EA1	:	A	CTTCTA	CTACTTTATTTTTT	CACACAACCTT	TCTA	ATG	TACCA	AAATTT	ACACATCTTCACATCTTCAA	CTTATTTT	AGCATATAA	AT	990
3EA2	:	A	CTTCTA	CTACTTTATTTTTT	CACACAACCTT	TCTA	ATG	TACCA	AAATTT	ACACATCTTCACATCTTCAA	CTTATTTT	AGCATATAA	AT	990

[illegible]

	2040	*	2060	*	2080	*	2100	*	2120	*	2140										
Koren	:											:									
Chao	:											:									
EU1	:											:									
EU2	:											:									
CAI1	:											:									
CAI2	:											:									
CAII1	:											:									
CAII2	:											:									
EA1	:											:									
EA2	:											:									
2Koren	:											:									
2Chao	:											:									
2EU1	:											:									
2EU2	:											:									
2CAI1,2	:											:									
2CAII1,2	:											:									
2EA1	:											:									
2EA2	:											:									
3Koren	:	AACAACTCAA	AAA	A	CTTTGCTCT	CT	CTT	CTTTAA	AA	ACAT	CACAAG	TG	AAACCTACCT	AA	CTT	CAAAATT	CAA	AA	ATCCCT	ATTCCA	: 352
3Chao	:	AACAACTCAA	AAA	A	CTTTGCTCT	CT	CTT	CTTTAA	AA	ACAT	CACAAG	TG	AAACCTACCT	AA	CTT	CAAAATT	CAA	AA	ATCCCT	ATTCCA	: 352
3EU1	:	AACAACTCAA	AAA	A	CTTTGCTCT	CT	CTT	CTTTAA	AA	ACAT	CACAAG	TG	AAACCTACCT	AA	CTT	CAAAATT	CAA	AA	ATCCCT	ATTCCA	: 917
3EU2	:	AACAACTCAA	AAA	A	CTTTGCTCT	CT	CTT	CTTTAA	AA	ACAT	CACAAG	TG	AAACCTACCT	AA	CTT	CAAAATT	CAA	AA	ATCCCT	ATTCCA	: 914
3CAI1	:	AACAACTCAA	AAA	A	CTTTGCTCT	CT	CTT	CTTTAA	AA	ACAT	CACAAG	TG	AAACCTACCT	AA	CTT	CAAAATT	CAA	AA	ATCCCT	ATTCCA	: 1176
3CAI2	:	AACAACTCAA	AAA	A	CTTTGCTCT	CT	CTT	CTTTAA	AA	ACAT	CACAAG	TG	AAACCTACCT	AA	CTT	CAAAATT	CAA	AA	ATCCCT	ATTCCA	: 1177
3CAII1	:	AACAACTCAA	AAA	A	CTTTGCTCT	CT	CTT	CTTTAA	AA	ACAT	CACAAG	TG	AAACCTACCT	AA	CTT	CAAAATT	CAA	AA	ATCCCT	ATTCCA	: 1180
3CAII2	:	AACAACTCAA	AAA	A	CTTTGCTCT	CT	CTT	CTTTAA	AA	ACAT	CACAAG	TG	AAACCTACCT	AA	CTT	CAAAATT	CAA	AA	ATCCCT	ATTCCA	: 1179
3EA1	:	AACAACTCAA	AAA	A	CTTTGCTCT	CT	CTT	CTTTAA	AA	ACAT	CACAAG	TG	AAACCTACCT	AA	CTT	CAAAATT	CAA	AA	ATCCCT	ATTCCA	: 1204
3EA2	:	AACAACTCAA	AAA	A	CTTTGCTCT	CT	CTT	CTTTAA	AA	ACAT	CACAAG	TG	AAACCTACCT	AA	CTT	CAAAATT	CAA	AA	ATCCCT	ATTCCA	: 1204

Alignment A. 2(cont.)

		*	2160	*	2180	*	2200	*	2220	*	2240		
Koren	:	-----		-----		-----		-----		-----		:	-
Chao	:	-----		-----		-----		-----		-----		:	-
EU1	:	-----		-----		-----		-----		-----		:	-
EU2	:	-----		-----		-----		-----		-----		:	-
CAI1	:	-----		-----		-----		-----		-----		:	-
CAI2	:	-----		-----		-----		-----		-----		:	-
CAI11	:	-----		-----		-----		-----		-----		:	-
CAI12	:	-----		-----		-----		-----		-----		:	-
EA1	:	-----		-----		-----		-----		-----		:	-
EA2	:	-----		-----		-----		-----		-----		:	-
2Koren	:	-----		-----		-----		-----		-----		:	-
2Chao	:	-----		-----		-----		-----		-----		:	-
2EU1	:	-----		-----		-----		-----		-----		:	-
2EU2	:	-----		-----		-----		-----		-----		:	-
2CAI1,2	:	-----		-----		-----		-----		-----		:	-
2CAI11,2	:	-----		-----		-----		-----		-----		:	-
2EA1	:	-----		-----		-----		-----		-----		:	-
2EA2	:	-----		-----		-----		-----		-----		:	-
3Koren	:	ACTCACCCCTAAGATGACCGGTATTCTTAACAATCCCTAACACATTGCTTGCTGCTCAAAATATAACCACTTTAC	:	438									
3Chao	:	ACTCACCCCTAAGATGCGCTTTATTCTATTCAGTCCTCTCAAAATCTAAACTATTTTAAATTCCTTAAAACTTCCTTAAAT	:	436									
3EU1	:	ACTCACCCCTAAGATGCGCTTTATTCTATTCAGTCCTCTCAAAATCTGAAGACTATTTTAAATTCCTTAAAACTTCCTTAAAT	:	1003									
3EU2	:	ACTCACCCCTAAGATGCGCTTTATTCTATTCAGTCCTCTCAAAATCTAAACTATTTTAAATTCCTTAAAACTTCCTTAAAT	:	1000									
3CAI1	:	ACTCACCCCTAAGATGCGCTTTATTCTATTCAGTCCTCTCAAAATCTAAACTATTTTAAATTCCTTAAAACTTCCTTAAAT	:	1262									
3CAI2	:	ACTCACCCCTAAGATGCGCTTTATTCTATTCAGTCCTCTCAAAATCTAAACTATTTTAAATTCCTTAAAACTTCCTTAAAT	:	1263									
3CAI11	:	ACTCACCCCTAAGATGCGCTTTATTCTATTCAGTCCTCTCAAAATCTAAACTATTTTAAATTCCTTAAAACTTCCTTAAAT	:	1266									
3CAI12	:	ACTCACCCCTAAGATGCGCTTTATTCTATTCAGTCCTCTCAAAATCTAAACTATTTTAAATTCCTTAAAACTTCCTTAAAT	:	1265									
3EA1	:	ACTCACCCCTAAGATGCGCTTTATTCTATTCAGTCCTCTCAAAATCTAAACTATTTTAAATTCCTTAAAACTTCCTTAAAT	:	1290									
3EA2	:	ACTCACCCCTAAGATGCGCTTTATTCTATTCAGTCCTCTCAAAATCTAAACTATTTTAAATTCCTTAAAACTTCCTTAAAT	:	1290									
	:	-----										:	-
		*	2260	*	2280	*	2300	*	2320	*	2340	*	
Koren	:	-----		-----		-----		-----		-----		-----	:
Chao	:	-----		-----		-----		-----		-----		-----	:
EU1	:	-----		-----		-----		-----		-----		-----	:
EU2	:	-----		-----		-----		-----		-----		-----	:
CAI1	:	-----		-----		-----		-----		-----		-----	:
CAI2	:	-----		-----		-----		-----		-----		-----	:
CAI11	:	-----		-----		-----		-----		-----		-----	:
CAI12	:	-----		-----		-----		-----		-----		-----	:
EA1	:	-----		-----		-----		-----		-----		-----	:
EA2	:	-----		-----		-----		-----		-----		-----	:
2Koren	:	-----		-----		-----		-----		-----		-----	:
2Chao	:	-----		-----		-----		-----		-----		-----	:
2EU1	:	-----		-----		-----		-----		-----		-----	:
2EU2	:	-----		-----		-----		-----		-----		-----	:
2CAI1,2	:	-----		-----		-----		-----		-----		-----	:
2CAI11,2	:	-----		-----		-----		-----		-----		-----	:
2EA1	:	-----		-----		-----		-----		-----		-----	:
2EA2	:	-----		-----		-----		-----		-----		-----	:
3Koren	:	-----AGTCTCTCTTATATGTCAGCAAAATTCGAACCA--CATGCTAGTCTGCTTTTCTTTTTCGCTCCCATATTTAAACATTA-----CCTATTATAT	:	532									
3Chao	:	AAATTATGATCTATCTTTATATATCAAAATTTCAACCAACATGCTAGCCCTTCCTTTCTTTTTCCTCCTATATTTAAACCTTATCTAGCACTATTATAT	:	542									
3EU1	:	AAATTATGATCTATCTTTATATATCAAAATTTCAACCAACATGCTAGCCCTTCCTTTCTTTTTCCTCCTATATTTAAACCTTATCTAGCACTATTATAT	:	1109									
3EU2	:	AAATTATGATCTATCTTTATATATCAAAATTTCAACCAACATGCTAGCCCTTCCTTTCTTTTTCCTCCTATATTTAAACCTTATCTAGCACTATTATAT	:	1106									
3CAI1	:	AAATTATGATCTATCTTTATATATCAAAATTTCAACCAACATGCTAGCCCTTCCTTTCTTTTTCCTCCTATATTTAAACCTTATCTAGCACTATTATAT	:	1368									
3CAI2	:	AAATTATGATCTATCTTTATATATCAAAATTTCAACCAACATGCTAGCCCTTCCTTTCTTTTTCCTCCTATATTTAAACCTTATCTAGCACTATTATAT	:	1369									
3CAI11	:	AAATTATGATCTATCTTTATATATCAAAATTTCAACCAACATGCTAGCCCTTCCTTTCTTTTTCCTCCTATATTTAAACCTTATCTAGCACTATTATAT	:	1372									
3CAI12	:	AAATTATGATCTATCTTTATATATCAAAATTTCAACCAACATGCTAGCCCTTCCTTTCTTTTTCCTCCTATATTTAAACCTTATCTAGCACTATTATAT	:	1371									
3EA1	:	AAATTATGATCTATCTTTATATATCAAAATTTCAACCAACATGCTAGCCCTTCCTTTCTTTTTCCTCCTATATTTAAACCTTATCTAGCACTATTATAT	:	1396									
3EA2	:	AAATTATGATCTATCTTTATATATCAAAATTTCAACCAACATGCTAGCCCTTCCTTTCTTTTTCCTCCTATATTTAAACCTTATCTAGCACTATTATAT	:	1396									
	:	-----											:
		2360	*	2380	*	2400	*	2420	*	2440	*		
Koren	:	-----		-----		-----		-----		-----		-----	:
Chao	:	-----		-----		-----		-----		-----		-----	:
EU1	:	-----		-----		-----		-----		-----		-----	:
EU2	:	-----		-----		-----		-----		-----		-----	:
CAI1	:	-----		-----		-----		-----		-----		-----	:
CAI2	:	-----		-----		-----		-----		-----		-----	:
CAI11	:	-----		-----		-----		-----		-----		-----	:
CAI12	:	-----		-----		-----		-----		-----		-----	:
EA1	:	-----		-----		-----		-----		-----		-----	:
EA2	:	-----		-----		-----		-----		-----		-----	:
2Koren	:	-----		-----		-----		-----		-----		-----	:
2Chao	:	-----		-----		-----		-----		-----		-----	:
2EU1	:	-----		-----		-----		-----		-----		-----	:
2EU2	:	-----		-----		-----		-----		-----		-----	:
2CAI1,2	:	-----		-----		-----		-----		-----		-----	:
2CAI11,2	:	-----		-----		-----		-----		-----		-----	:
2EA1	:	-----		-----		-----		-----		-----		-----	:
2EA2	:	-----		-----		-----		-----		-----		-----	:
3Koren	:	TATTCTTCTCATTAAGATGCTCATAAATTTAAAAAGTTTCCTTAAAA--CTAAGATATATCTTAAAGATTTCACATTTCTAA--CAA	:	627									
3Chao	:	TATTCTTCTCATTAAGAAAT--CTCTTAAATTTAAAGCTTTTCCTTAAAA--CTAAGATATATCTTAAAGATTTCACATTTCTTAA--CAA	:	632									
3EU1	:	TATTCTTCTCATTAAGAAAT--CTCTTAAATTTAAAGCTTTTCCTTAAAAACATTAAGATATATCTTAAAGATTTCACATTTCTTAA--CAA	:	1200									
3EU2	:	TATTCTTCTCATTAAGAAAT--CTCTTAAATTTAAAGCTTTTCCTTAAAAACATTAAGATATATCTTAAAGATTTCACATTTCTTAA--CAA	:	1197									
3CAI1	:	TATTCTTCTCATTAAGAAAT--CTCTTAAATTTAAAGCTTTTCCTTAAAA--CTAAGATATATCTTAAAGATTTCACATTTCTTAA--CAA	:	1458									
3CAI2	:	TATTCTTCTCATTAAGAAAT--CTCTTAAATTTAAAGCTTTTCCTTAAAA--CTAAGATATATCTTAAAGATTTCACATTTCTTAA--CAA	:	1454									
3CAI11	:	TATTCTTCTCATTAAGAAAT--CTCTTAAATTTAAAGCTTTTCCTTAAAA--CTAAGATATATCTTAAAGATTTCACATTTCTTAA--CAA	:	1462									
3CAI12	:	TATTCTTCTCATTAAGAAAT--CTCTTAAATTTAAAGCTTTTCCTTAAAAACATTAAGATATATCTTAAAGATTTCACATTTCTTAA--CAA	:	1462									
3EA1	:	TATTCTTCTCATTAAGAAAT--CTCTTAAATTTAAAGCTTTTCCTTAAAAACATTAAGATATATCTTAAAGATTTCACATTTCTTAA--CAA	:	1487									
3EA2	:	TATTCTTCTCATTAAGAAAT--CTCTTAAATTTAAAGCTTTTCCTTAAAAACATTAAGATATATCTTAAAGATTTCACATTTCTTAA--CAA	:	1487									
	:	-----											:

Alignment A. 5 Alignment of amino acid sequences of mature proteins from *GH I* and *GH II* genes. GHI - mature protein sequence of the *GH I* gene; GHISUB - mature protein sequence of the *GH I* gene with substitution from phenylalanine to isoleucine Phe/Ile at position 84 of mature protein; GHII - mature protein sequence of the *GH II* gene.

GHI	:	SDN	Q	E	F	N	N	A	V	I	R	V	G	H	L	H	I	A	A	M	N	D	E	E	D	S	L	L	E	E	R	R	D	I	S	I	F	F	I	S	F	C	N	S	D	I	E	A	P	A	I	D	E	T	P	C	S	S	M	L	R	L	L	I	R	I	S	F	R	L	I	E	S	W	E	F	S	Q	S	L	S	I	S	T	V	S	N	S	L	T	V	G	N	P	N	Q	6	T	E	K	L	A	D	L	:	112				
GHISub	:	SDN	Q	E	F	N	N	A	V	I	R	V	G	H	L	H	I	A	A	M	N	D	E	E	D	S	L	L	E	E	R	R	D	I	S	I	F	F	I	S	F	C	N	S	D	I	E	A	P	A	I	D	E	T	P	C	S	S	M	L	R	L	L	I	R	I	S	F	R	L	I	E	S	W	E	I	S	Q	S	L	S	I	S	T	V	S	N	S	L	T	V	G	N	P	N	Q	6	T	E	K	L	A	D	L	:	112				
GHII	:	SDN	Q	E	F	N	N	A	V	I	R	V	G	H	L	H	I	A	A	M	N	D	E	E	D	S	L	L	E	E	R	R	D	I	S	I	F	F	I	S	F	C	N	S	D	I	E	A	P	A	I	D	E	T	P	C	S	S	M	L	R	L	L	I	R	I	S	F	R	L	I	E	S	W	E	F	S	Q	T	I	S	T	V	S	N	S	L	T	V	G	N	P	N	Q	6	T	E	K	L	A	D	L	:	112						
		SDN	Q	R	L	F	N	N	A	V	I	R	V	Q	H	L	H	Q	L	A	A	K	M	N	D	F	E	D	S	L	L	P	E	R	R	Q	L	S	K	I	F	P	L	S	F	C	N	S	D	I	E	A	P	a	G	K	D	E	T	Q	K	S	S	M	L	K	L	I	R	I	S	F	R	L	I	E	S	W	E	f	P	S	Q	3	L	S	G	T	V	S	N	S	L	T	V	G	N	P	N	Q	6	T	E	K	L	A	D	L		
GHI	:	M	E	I	S	V	L	I	Q	A	C	L	D	S	E	N	M	D	N	D	S	L	E	F	I	L	T	M	C	E	N	N	I	R	S	F	E	L	I	A	C	F	K	D	M	H	I	V	E	T	L	V	A	N	C	R	R	S	L	D	S	N	C	T	L	*	:	188																																										
GHISub	:	M	E	I	S	V	L	I	Q	A	C	L	D	S	E	N	M	D	N	D	S	L	E	F	I	L	T	M	C	E	N	N	I	R	S	F	E	L	I	A	C	F	K	D	M	H	I	V	E	T	L	V	A	N	C	R	R	S	L	D	S	N	C	T	L	*	:	188																																										
GHII	:	M	E	I	S	V	L	I	K	G	C	L	D	S	E	N	M	D	N	D	S	L	E	F	I	L	T	M	C	E	N	N	I	R	S	F	E	L	I	A	C	F	K	D	M	H	I	V	E	T	L	V	A	N	C	R	R	S	L	D	S	N	C	T	L	*	:	188																																										
		K	M	G	I	S	V	L	I	q	a	C	L	D	G	Q	P	N	M	D	N	D	S	L	P	F	E	D	F	Y	L	T	M	G	E	N	N	L	R	E	S	F	R	L	L	A	C	F	K	K	D	M	H	K	V	E	T	Y	L	R	V	A	N	C	R	R	S	L	D	S	N	C	T	L																																				

Table A. 1 Polymorphisms in common carp *GH I* gene. **Number** - number of polymorphic sites. **bp** - site on the Alignment A.I. EU - sample from Europe (River Rhine), CA I, CA II - samples from Central Asia (Uzbekistan), EA - sample from East Asia (China) (both clones were identical)

Intron I									
Number	1	2	3	4	5	6	Length of Intron bp		Content %
bp	49	186	187	188	195	204	C+G T+A		
Sample									
EU clone 1	T	A	A	A	A	-	284	25.4	74.6
EU clone 2	T	A	A	A	-	-	283	25.4	74.6
CA I clone 1	A	A	-	-	A	A	283	25.4	74.6
CA I clone 2	A	-	-	-	A	A	282	25.5	74.5
CA II clone 1	T	A	A	-	A	-	283	25.4	74.6
CA II clone 2	T	A	A	-	A	A	284	25.4	74.6
EA clones 1 and 2	T	A	A	-	A	-	283	25.4	74.6

Table A 1 (continued)

Exon II		Intron II	
Number	7	8	Content %
bp	345	657	Length of Intron bp C+G T+A
GTG-GTA			
Val-Val			
Sample			
EU clone 1	G	A	228 26.7 73.3
EU clone 2	G	A	228 26.7 73.3
CA I clone 1	A	G	228 27.2 72.8
CA I clone 2	A	G	228 27.2 72.8
CA II clone 1	G	G	228 27.2 72.8
CA II clone 2	A	G	228 27.2 72.8
EA clones 1, 2	G	G	228 27.2 72.8

Table A1 (continued)

Intron III																
Number	9	10	11	12	13	14	15	16	17	18	19	20	21	Length of Intron bp	Content %	
bp	805	861	895	944	955	972-1312	1029	1082	1117	1123	1187	1240	1358-1361		C+G T+A	
Sample	deletion										deletion					
EU clone 1	T	A	T	G	T		G	A	A	C	G	A	- 4 bp	569	33 67	
EU clone 2	T	A	T	G	T		G	A	A	C	G	A	- 4 bp	569	33 67	
CA I clone 1	A	G	G	A	T		A	T	T	G	A	C	CTTC	573	33.2 66.8	
CA I clone 2	A	G	G	A	T		A	T	T	G	A	C	CTTC	573	33.2 66.8	
CA II clone 1	A	G	G	A	C	-341bp	-	-	-	-	-	-	CTTC	232	37.5 62.5	
CA II clone 2	A	G	G	A	T		A	T	T	G	A	C	CTTC	573	33.2 66.8	
EA clones 1, 2	A	G	G	A	C	-341bp	-	-	-	-	-	-	CTTC	232	37.5 62.5	

Table A1 (continued)

Exon IV				Intron IV			
Number	22	23	24	25	26	27	Content %
bp	1378	1412	1417	1585	1597	1608	C+G T+A
	CTT-CTA	TTC-ATC	CCT-CCA				
	Leu-Leu	Phe-Ile	Pro-Pro				
Sample							
EU clone 1	T	T	T	T	C	C	42.5 57.5
EU clone 2	T	T	T	T	C	C	42.5 57.5
CA I clone 1	T	T	A	T	C	C	42.5 57.5
CA I clone 2	T	T	A	T	C	C	42.5 57.5
CA II clone 1	A	T	A	C	A	A	41.5 58.5
CA II clone 2	T	A	A	T	C	C	42.5 57.5
EA clones 1, 2	A	T	A	C	A	A	41.5 58.5

Table A1 (continued)

	Exon V		3' End	
Number	28	29	30	31
bp	1667	1893	1936	2015
Sample	AAC-AAT			
	Asn-Asn			
	EU clone 1	C	A	C
	EU clone 2	C	A	C
	CA I clone 1	C	G	A
	CA I clone 2	C	G	A
	CA II clone 1	T	A	C
	CA II clone 2	T	A	C
	EA clones 1, 2	T	A	C

Table A. 2 Polymorphisms in common carp *GH II* gene. **Number** - number of polymorphic sites. **bp** - site on the Alignment A.2. EU - sample from Europe (River Rhine), CA I, CA II - samples from Central Asia (Uzbekistan), EA - sample from East Asia (China)

<i>GH-a</i> + <i>GH-k</i> primer pair, fragment from first to second exon												
Intron I												
Number	1	2	3	4	5	6	7	8	9	10	Length of Intron bp	
bp	29	91	117	189	224	227	263	312	329	358	C+G	T+A
Sample	Micro-satellite		Microsatellite		Microsatellite		Microsatellite		Deletion			
EU clone 1	T	(CA) ₈	(TA) ₆ TG(TA) ₅ TC(TA) ₆ GATCA	(TA) ₅ T ₇ CT ₆	T	(TTA) ₃ T ₁₂	A ₁₂ TTCAT ₃	A		T	355	76.9
EU clone 2	T	(CA) ₈	(TA) ₁₄ TC(TA) ₆ GATCA,	(TA) ₅ T ₇ CT ₆	T	(TTA) ₃ T ₁₃	A ₁₂ TTCAT ₃	A		T	359	77.4
CA I clone 1	G	CA	(TA) ₁₉ TCA	(TA) ₃ T ₁₁ CT ₇	G	(TTA) ₅ T ₁₂	A ₄ T ₁₀	C		-	337	78
CA I clone 2	G	CA	(TA) ₂₀ TCA	(TA) ₃ T ₁₁ CT ₇	G	(TTA) ₅ T ₁₂	A ₄ T ₁₀	C		-	340	78.2
CA II clone 1	G	CA	(TA) ₁₈ TCA	(TA) ₅ T ₈ CT ₇ A	G	(TTA) ₅ T ₁₂	A ₄ T ₁₀	C		-	338	77.8
CA II clone 2	G	CA	(TA) ₁₉ TCA	(TA) ₅ T ₈ CT ₇ A	G	(TTA) ₅ T ₁₂	A ₄ T ₁₀	C		-	340	78.04
EA clone 1	G	(CA) ₁₃	(TA) ₃₂ TTA	(TA) ₅ T ₆ CT ₇	G	(TTA) ₃₂ T ₁₇	A ₄ T ₉	A	-49	-	342	79.2
EA clone 2	G	(CA) ₁₃	(TA) ₃₂ A	(TA) ₅ T ₆ CT ₇	G	(TTA) ₇ T ₁₀	A ₄ T ₉	A	-49	-	342	79.2

Table A2 (continued)

<i>GH-g + GH-h primer pair, fragment from second exon to third intron¹</i>						
		Intron II		Exon III	Intron III	
Number	11	Length of Intron bp	Content %		12	13
bp	740		C+G	T+A	893	940
						941
Sample					AAA-AAAG Lys-Lys	
2 EU clone 1	A	235	31.5	68.5	A	G
2 EU clone 2	T	235	31.5	69.5	A	G
2 CA I clones 1, 2	T	235	30.6	69.4	A	G
2 CA II clones 1, 2	T	235	30.6	69.4	A	G
2 EA clone 1	T	235	30.6	69.4	A	T
2 EA clone 2	T	235	30.6	69.4	G	T

¹Both clones from CA I and both clones from CA II were identical

Table A2 (continued)

GH-e + GH-b primer pair, fragment from third intron to non-coding 3' end																							
Intron III																							
Number		15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32				
bp		972	973	978	1005	1027	1034	1044	1089	1104	1112-	1194	1241	1270	1324	1325	1326	1349	1381				
Sample												Deletion											Deletion
3 EU clone 1	T	G	-	C	A	A	A	T	G	T	-284	-	-	-	-	-	-	-	-	-			
3 EU clone 2	T	G	-	C	A	A	A	T	G	T	-284	-	-	-	-	-	-	-	-	-			
3 CA I clone 1	C	A	A	A	T	T	G	A	A	A		A	T	-17	A	A	-	A	-	-			
3 CA I clone 2	C	A	A	A	T	T	G	A	A	A		A	T	-17	A	A	-	A	-	-			
3 CA II clone 1	C	A	A	A	T	T	G	A	A	A		A	T	-17	A	A	A	A	-	-			
3 CA II clone 2	C	A	A	A	T	T	G	A	A	A		A	T	-17	A	A	A	A	-	-			
3 EA clone 1	C	A	A	A	C	T	G	A	A	A		T	A		A	A	T	T	A	A			
3 EA clone 2	C	A	A	A	C	T	G	A	A	A		T	A		-	-	T	T	A	A			

Table A2 (continued)

Intron III										Exon IV		
Number	33	34	35	36	37	38	39	Content %		Length of Intron bp		
bp	1400	1420	1488	1542	1545	1546	1641	C+G	T+A			
Sample											GCC-GCT	
3 EU clone 1	A	A	A	T	T	T ₁₅	TTC	28.1	71.9	441	C	
3 EU clone 2	A	A	A	T	T	T ₁₃	TTC	28.3	71.7	439	C	
3 CA I clone 1	G	T	G	C	T	T ₁₂	-	24.7	75.3	701	T	
3 CA I clone 2	G	T	G	C	T	T ₁₃	-	24.6	75.4	702	T	
3 CA II clone 1	G	T	G	C	T	T ₁₅	-	24.5	75.5	705	T	
3 CA II clone 2	G	T	G	C	T	T ₁₄	-	24.6	75.3	704	C	
3 EA clone 1	G	T	G	C	G	T ₁₈	TTC	24	76	729	C	
3 EA clone 2	G	T	G	C	G	T ₂₀	TTC	24	76	729	C	

Table A2 (continued)

	<i>Intron IV</i>				<i>3' End</i>		
	41 1845	Length of Intron bp	Content %		42 2404	43 2405	44 2406
Number* bp			C+G	T+A			
Sample							
3 EU clone 1	T	143	40	60	A	C	A
3 EU clone 2	-	142	40.1	59.9	A	C	A
3 CA I clone 1	-	142	40.1	59.9	A	A	-
3 CA I clone 2	-	142	40.1	59.9	A	A	-
3 CA II clone 1	-	142	40.1	59.9	-	C	A
3 CA II clone 2	-	142	40.1	59.9	A	C	A
3 EA clone 1	-	142	40.1	59.9	A	C	A
3 EA clone 2	-	142	40.1	59.9	A	C	A

Table A. 3 Primary data of the *GH II/GH I* mRNA level ratios.

GH II/GH I mRNA level ratios							
Group (age or sex)	Sample ID	Number of fish	Replicate			Mean	STD
			a	b	c		
15 days	II-1	30 fry	0.72	0.80	0.82	0.78	0.05
	II-2	30	0.77	0.81	0.76	0.78	0.03
	II-3	30	0.76	0.82	0.74	0.77	0.04
	II-4	30	1.21	1.10	0.96	1.09	0.13
	II-5	30	1.03	1.00	1.06	1.03	0.03
	II-6	30	0.99	1.08	1.15	1.07	0.08
Mean value						0.92	0.16
22 days	IIIs-1	30 fry	0.79	0.72	0.88	0.80	0.08
	IIIs-2	30	0.91	0.93	0.72	0.85	0.12
	IIIs-3	30	0.69	0.68	0.84	0.74	0.09
	IIIs-4	30	0.99	1.11	0.97	1.02	0.08
	IIIs-5	30	1.07	1.09	1.18	1.11	0.06
	IIIs-6	30	0.98	1.07	1.13	1.06	0.08
Mean value						0.93	0.16
29 days	IIIr-1	30 fry	0.96	0.85	1.02	0.94	0.09
	IIIr-2	30	0.75	0.72	0.69	0.72	0.03
	IIIr-3	30	0.85	0.78	0.84	0.82	0.04
	IIIr-4	30	0.87	1.00	0.98	0.95	0.07
	IIIr-5	30	0.91	0.96	1.01	0.96	0.05
	IIIr-6	30	0.91	0.95	1.01	0.96	0.05
Mean value						0.89	0.10
47 days	IV-1	2 heads	0.56	0.58	0.58	0.57	0.01
	IV-2	3	0.84	0.90	0.91	0.88	0.04
	IV-3	2	1.04	1.06	1.03	1.04	0.02
	IV-4	3	0.97	0.96	0.82	0.92	0.08
	IV-5	3	0.91	1.03	1.09	1.01	0.09
	IV-6	3	0.93	0.94	1.02	0.96	0.05
Mean value						0.90	0.17
63 days	V-13	2 heads	0.75	0.72	0.76	0.74	0.02
	V-14	2	0.76	0.71	0.60	0.69	0.08
	V-20	2	1.06	1.07	1.05	1.06	0.01
	V-16	2	0.98	1.03	1.08	1.03	0.05
	V-17	2	1.08	1.00	0.98	1.02	0.05
	V-18	2	0.98	0.91	1.07	0.99	0.08
Mean value						0.92	0.16
85 days	VI-1	3 whole brains	0.59	0.50	0.58	0.56	0.05
	VI-2	3	0.59	0.56	0.56	0.57	0.02
	VI-3	3	1.29	1.21	1.23	1.24	0.04
	VI-4	3	0.69	0.73	0.77	0.73	0.04
	VI-5	3	0.76	0.83	0.80	0.80	0.04
	VI-6	3	0.84	0.86	0.93	0.88	0.05
Mean value						0.80	0.24

Table A.3 Continued

GH II/GH I mRNA level ratios							
Group (age or sex)	Sample ID	Number of fish	Replicate			Mean	STD
			a	b	c		
1 year	1_1	1 pituitary gland	0.60	0.51	0.51	0.54	<i>0.05</i>
	1_2	1	0.66	0.67	0.57	0.63	<i>0.06</i>
	1_3	1	0.50	0.54	0.46	0.50	<i>0.04</i>
	1_4	1	0.95	0.98	0.97	0.97	<i>0.02</i>
	1_5	1	0.61	0.64	0.64	0.63	<i>0.02</i>
	1_6	1	0.78	0.81	0.80	0.80	<i>0.02</i>
	1_7	1	0.87	0.80	0.79	0.82	<i>0.04</i>
	1_8	1	0.93	0.94	0.96	0.94	<i>0.02</i>
	1_9	1	0.97	0.92	0.86	0.92	<i>0.06</i>
	1_10	1	0.94	0.98	0.91	0.94	<i>0.04</i>
Mean value						0.77	<i>0.18</i>
3 years Female	3_1 F	1 pituitary gland	0.73	0.72	0.75	0.73	<i>0.02</i>
	3_2 F	1	0.41	0.40	0.42	0.41	<i>0.01</i>
	3_3 F	1	0.50	0.53	0.47	0.50	<i>0.03</i>
	3_4 F	1	0.77	0.81	0.75	0.78	<i>0.03</i>
	3_5 F	1	0.68	0.72	0.67	0.69	<i>0.03</i>
	3_6 F	1	0.69	0.61	0.64	0.65	<i>0.04</i>
	3_7 F	1	0.74	0.74	0.73	0.74	<i>0.01</i>
	3_8 F	1	0.83	0.82	0.86	0.84	<i>0.02</i>
	3_9 F	1	1.04	0.95	0.92	0.97	<i>0.06</i>
	3_10 F	1	0.83	0.87	0.93	0.88	<i>0.05</i>
Mean value						0.72	<i>0.16</i>
3 years Male	3_1 M	1 pituitary gland	0.42	0.51	0.39	0.44	<i>0.06</i>
	3_2 M	1	0.68	0.70	0.64	0.67	<i>0.03</i>
	3_3 M	1	0.39	0.33	0.27	0.33	<i>0.06</i>
	3_4 M	1	1.06	1.12	1.20	1.13	<i>0.07</i>
	3_5 M	1	0.93	0.97	0.90	0.93	<i>0.04</i>
	3_6 M	1	0.87	0.74	0.78	0.80	<i>0.07</i>
	3_7 M	1	0.81	0.80	0.80	0.80	<i>0.01</i>
	3_8 M	1	0.86	0.91	0.92	0.90	<i>0.03</i>
	3_9 M	1	0.77	0.71	0.79	0.76	<i>0.04</i>
Mean value						0.75	<i>0.24</i>
Cold adapted 10 months old fish		10 pituitary glands	0.32	0.29	0.33	0.31	

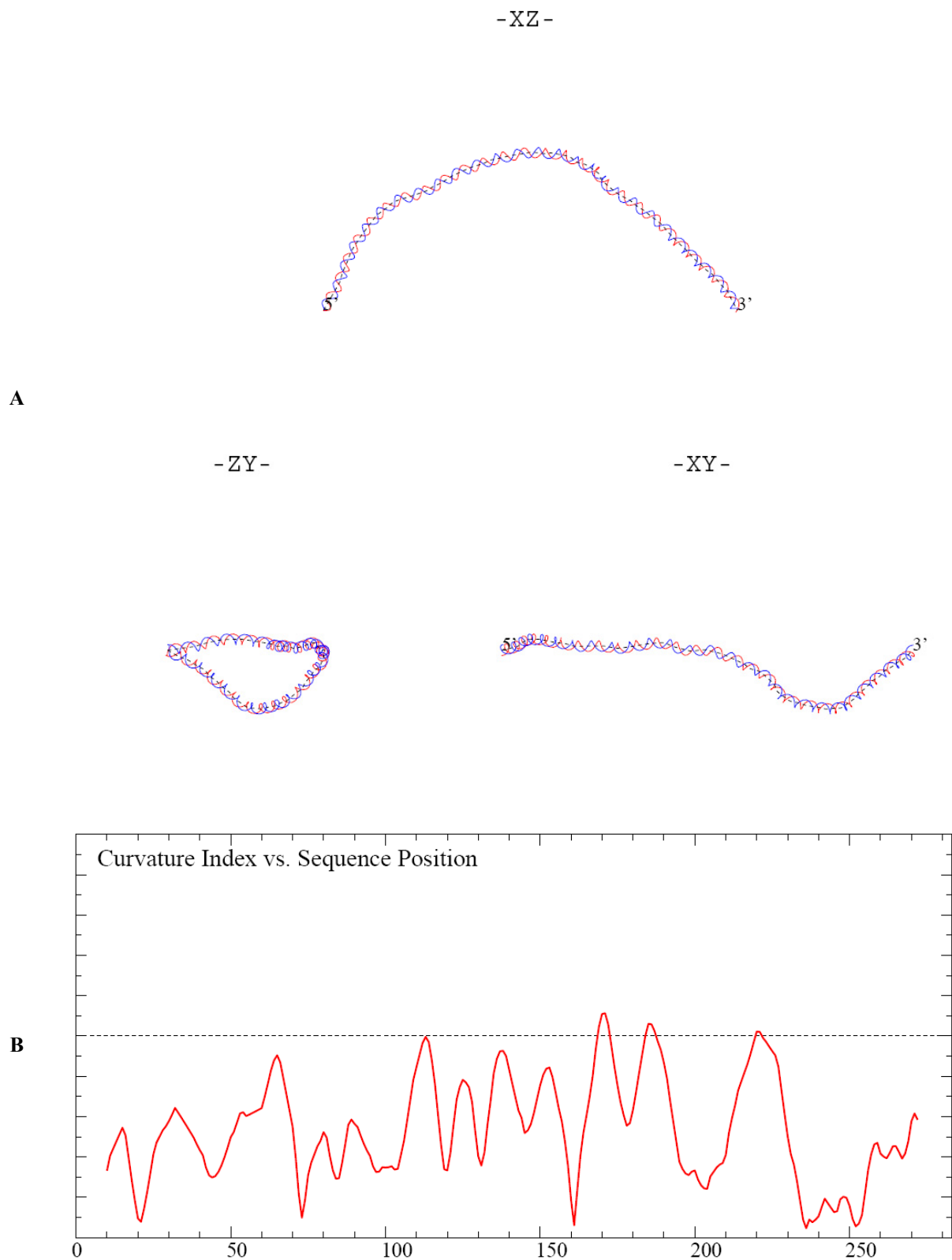


Figure A. 1 Preliminary predicted intrinsic DNA curvature in common carp *GH I* Intron I. A. DNA molecule projections on coordinate planes; B. Curvature index vs. Sequence position.

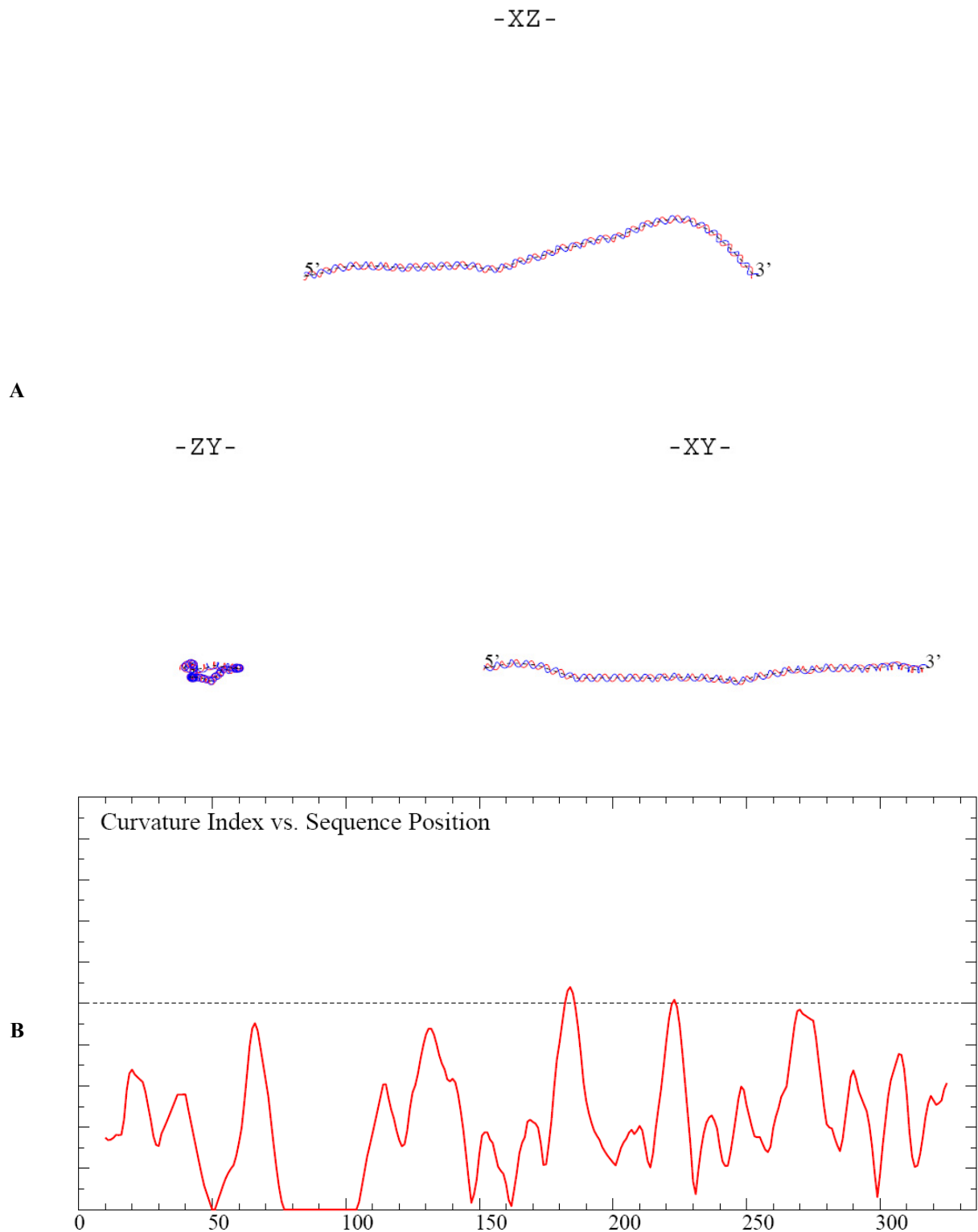
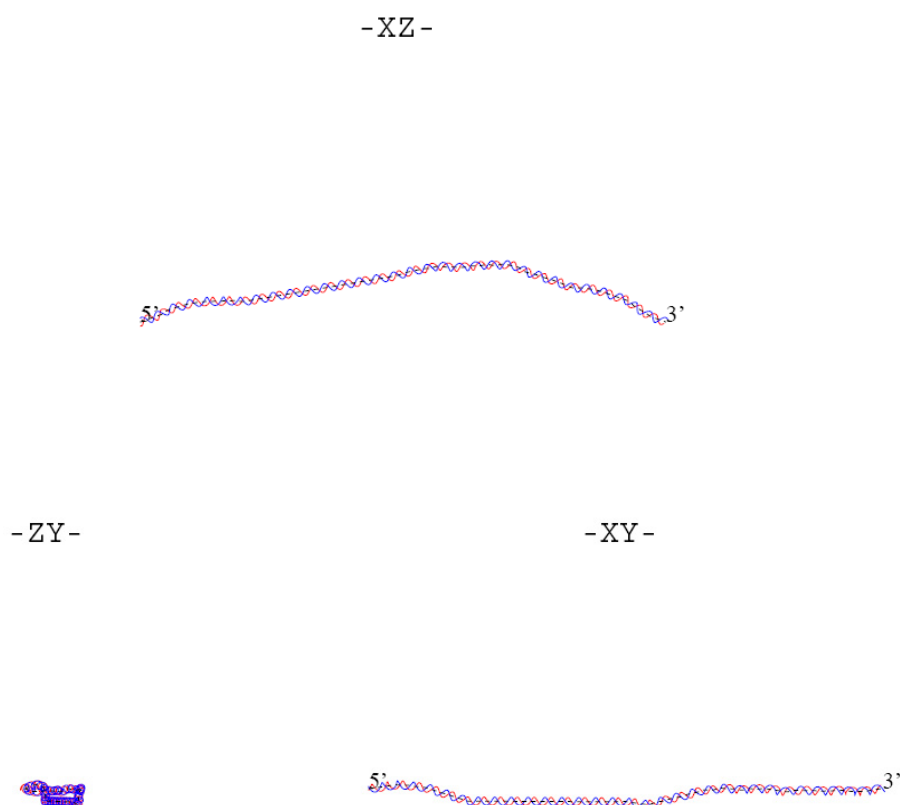


Figure A. 2 Preliminary predicted intrinsic DNA curvature in common carp *GH II* Intron I. Alleles with shortest poly-CA and poly-TA (Central Asia). The sequence position of poly-CA and poly-TA on the graph of Curvature index ranges from 80 to about 100 bp. A. DNA molecule projections on coordinate planes; B. Curvature index vs. Sequence position.

A



B

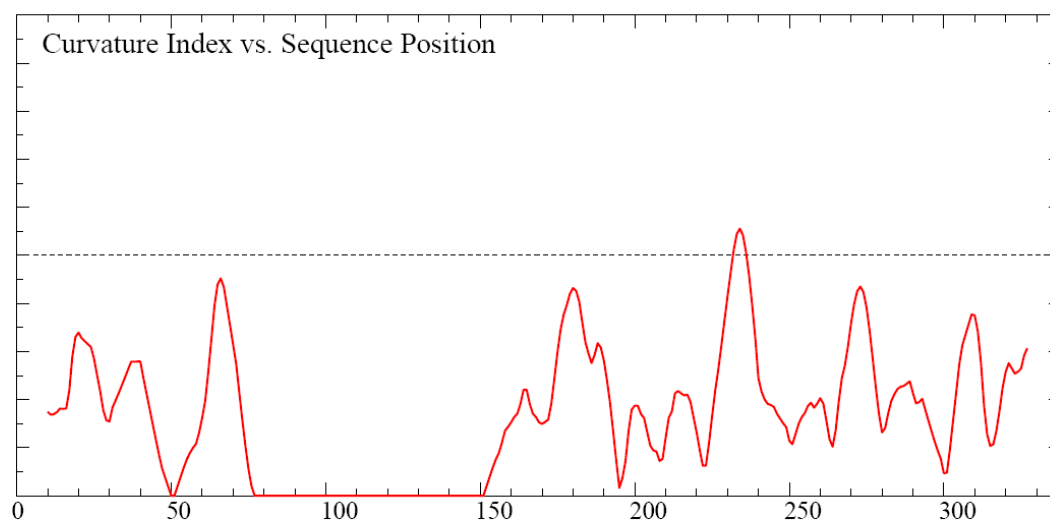


Figure A. 3Preliminary predicted intrinsic DNA curvature in common carp *GH II* Intron I. Alleles with longest poly-CA and poly-TA (China). The sequence position of poly-CA and poly-TA on the graph of Curvature index ranges from about 80 to 150 bp. A. DNA molecule projections on coordinate planes; B. Curvature index vs. Sequence position.

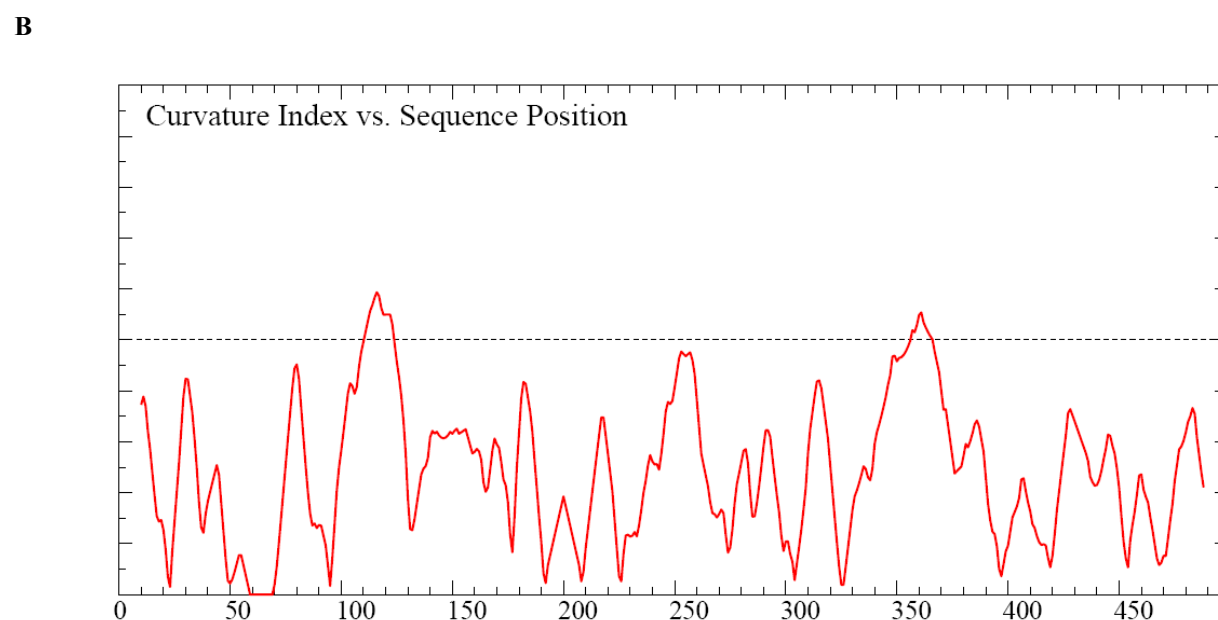
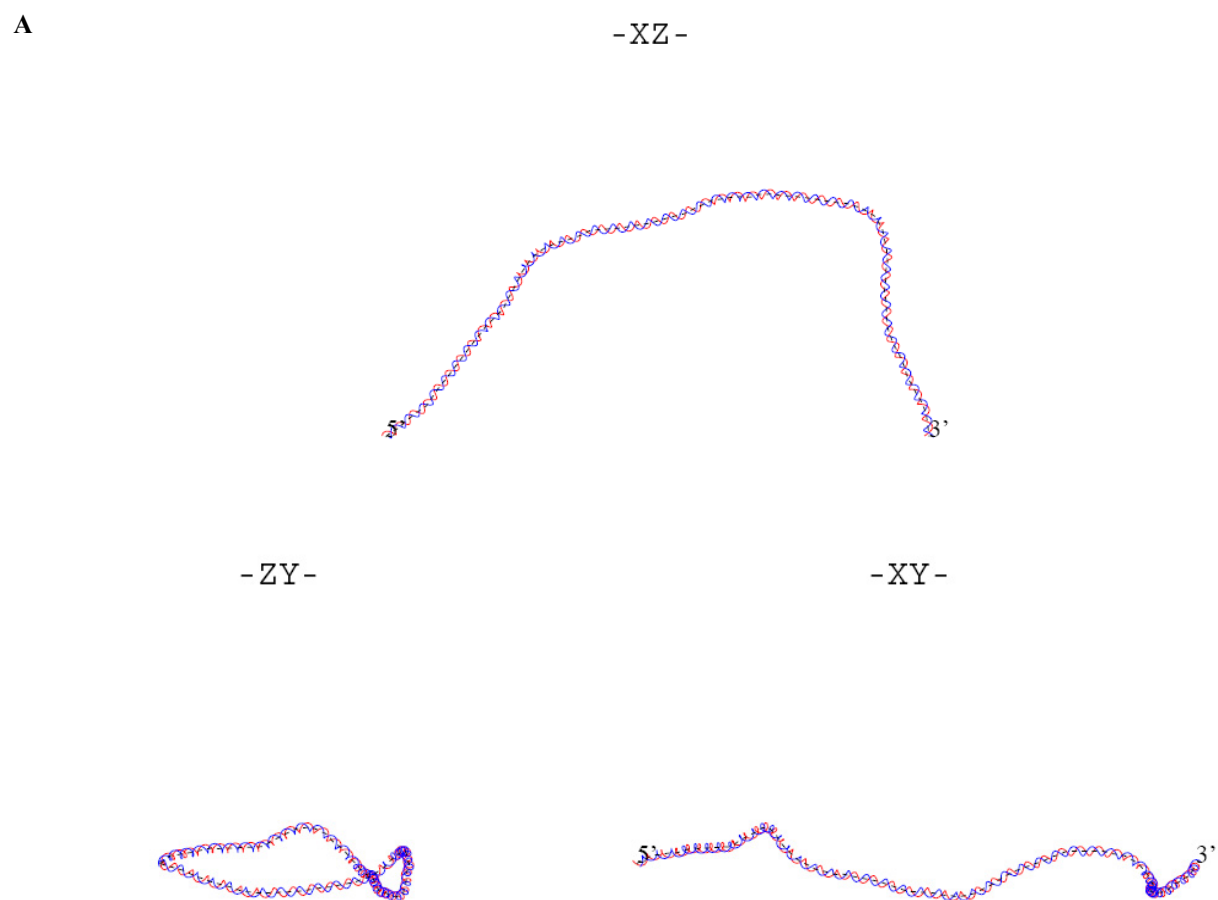
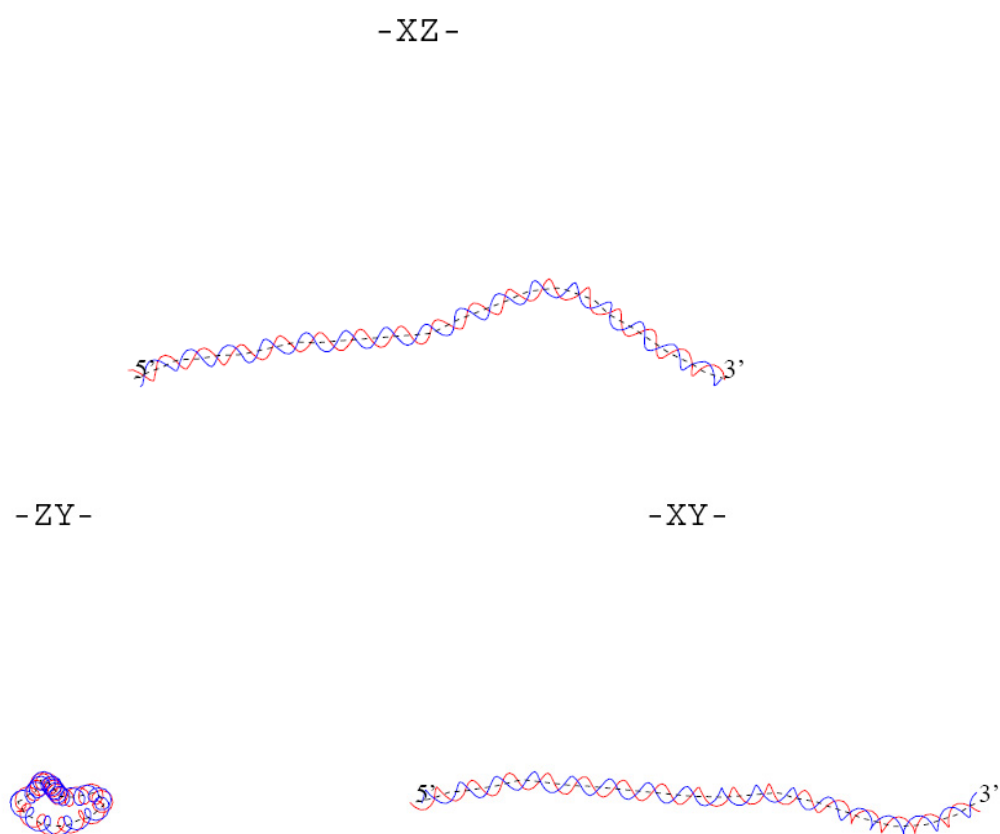


Figure A. 4 Preliminary predicted intrinsic DNA curvature in common carp *GH I* Intron III without deletion of 341 bp. A. DNA molecule projections on coordinate planes; B. Curvature index vs. Sequence position.

A



B

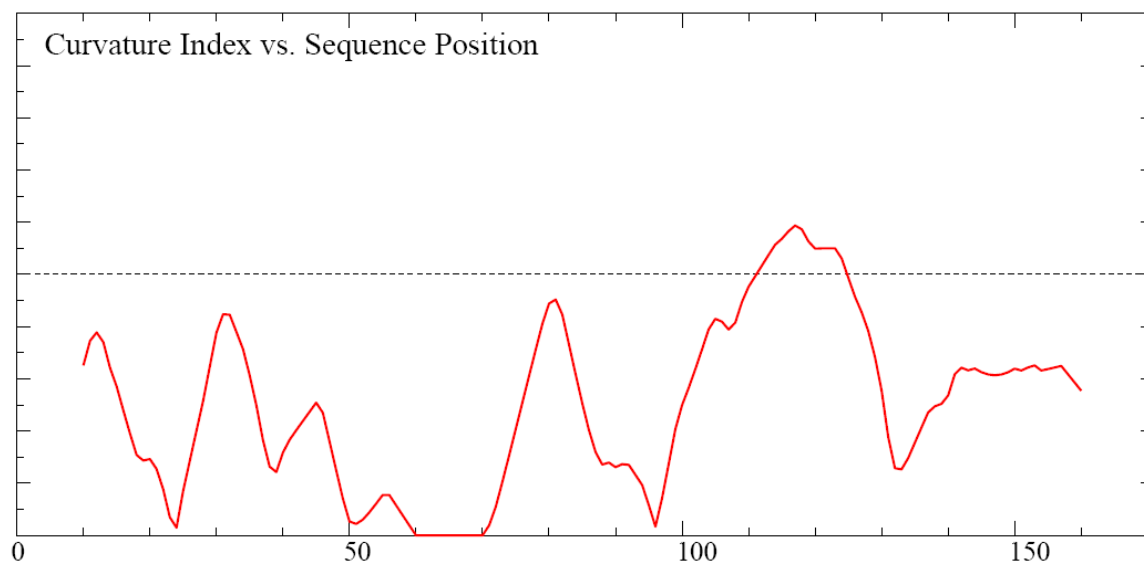


Figure A. 5 Preliminary predicted intrinsic DNA curvature in common carp *GH I* Intron III with deletion of 341 bp. A. DNA molecule projections on coordinate planes; B. Curvature index vs. Sequence position.

A



B

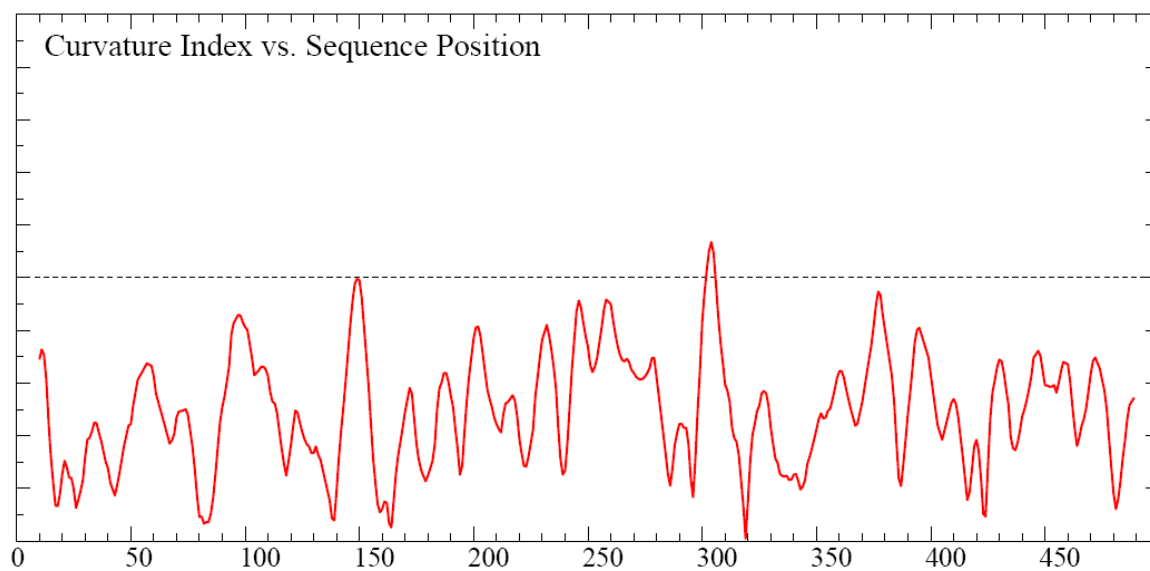


Figure A. 6 Preliminary predicted intrinsic DNA curvature in common carp *GH II* Intron III without deletion of 284 bp. A. DNA molecule projections on coordinate planes; B. Curvature index vs. Sequence position.

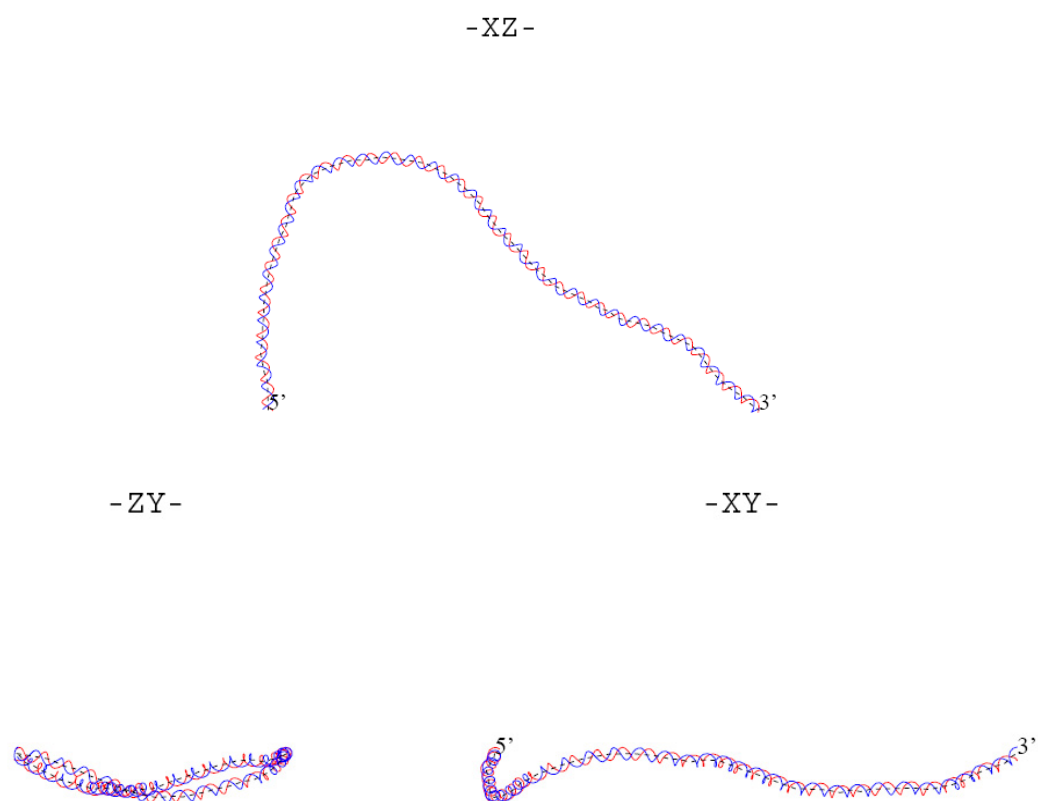
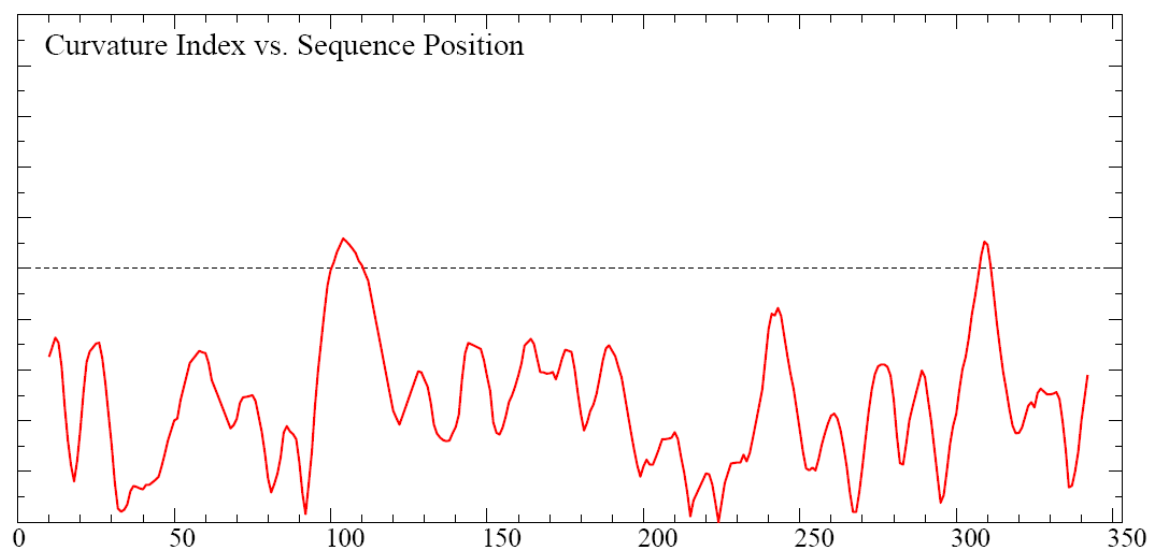
A**B**

Figure A. 7 Preliminary predicted intrinsic DNA curvature in common carp *GH II* Intron III with deletion of 284 bp. A. DNA molecule projections on coordinate planes; B. Curvature index vs. Sequence position.

Test of eight possible combinations of expression vectors and hosts to obtain recombinant protein in soluble fraction

Production of protein with the combination of expression host and expression vector that allows production of soluble recombinant common carp GH

Bacteria lysis and preparation of soluble fraction of proteins for loading on HiTrap column

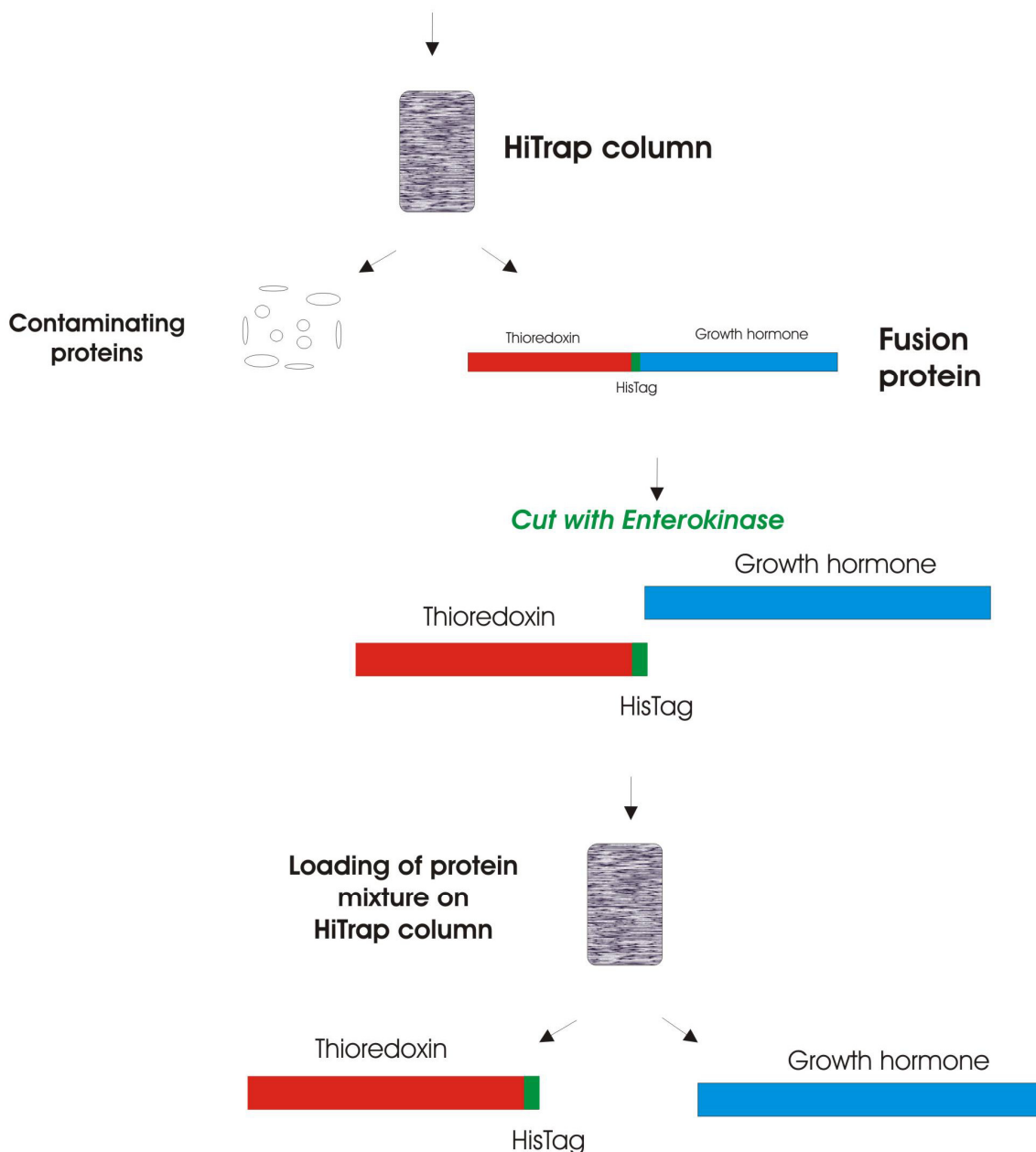


Figure A. 8 Scheme of recombinant cGH production and purification.

GenBank accession numbers of GH I and GH II sequences obtained in the present study

EU - sample from Europe (River Rhine),

CA I, CA II - samples from Central Asia (Uzbekistan),

EA - sample from East Asia (China).

GH I

Complete gene sequence

EU 1 EU333981

EU 2 EU333982

CA I 1 EU333983

CA I 2 EU333984

CA II 1 EU333985

CA II 2 EU333986

EA EU333987

GH II

GH-a + GH-k primer pair, fragment from first to second exon

EU 1 EU340116

EU 2 EU340117

CA I 1 EU340118

CA I 2 EU340119

CA II 1 EU340120

CA II 2 EU340121

EA 1 EU340122

EA 2 EU340123

GH-g + GH-h primer pair, fragment from second exon to third intron

2 EU 1 EU340124

2 EU 2 EU340125

2 CA I 1, 2	EU340126
2 CA II 1, 2	EU340127
2 EA 1	EU340128
2 EA 2	EU340129

GH-e + GH-b primer pair, fragment from third intron to non-coding 3' end

3 EU 1	EU340130
3 EU 2	EU340131
3 CA I 1	EU340132
3 CA I 2	EU340133
3 CA II 1	EU340134
3 CA II 2	EU340135
3 EA 1	EU340136
3 EA 2	EU340137

Abbreviations

ADP receptor	adenosine diphosphate receptor
BSA	bovine serum albumin
CA1	first fish from wild Lake Tuzkan population, Uzbekistan
CA2	second fish from wild Lake Tuzkan population, Uzbekistan
CAM 1 and CAM 2	duplicated c-myc genes in common carp
c-myc	proto-oncogene
CMC	calcium-magnesium containing medium
CMF	calcium-magnesium-free medium
CNS	central nervous system
DEPC water	diethylpyrocarbonate treated water
dN	number of non-synonymous nucleotide substitutions per non-synonymous site
dS	number of synonymous nucleotide substitutions per synonymous site
E2	estradiol
EA	fish from domesticated Wuhan population, China
ECP	eosinophil cationic protein
EDN	eosinophil-derived neurotoxin
EU	fish from wild/feral River Rhine population, Germany
GH	growth hormone
GH	growth hormone gene
GHR	growth hormone receptor
GHR	growth hormone receptor gene
GRF	growth hormone releasing factor
GRH	gonadotropin-releasing hormone
GRH	gonadotropin-releasing hormone gene

GTH	gonadotropin
hGH	human growth hormone
HisTag	peptide of six histidine residues fused to recombinant protein in expression vector
IGF-I	insulin-like growth factor I
IGF-I	insulin-like growth factor I gene
IGF-II	insulin-like growth factor II gene
IMAC	immobilized metal-ion affinity chromatography
IPTG	Isopropyl β -D-1-thiogalactopyranoside
LB	Luria-Bertani medium
MEM	minimal essential medium
PCR	polymerase chain reaction
Pit-1	pituitary-specific transcription factor
Pit-1	pituitary-specific transcription factor gene
pNC	proportion of conservative non-synonymous differences per conservative non-synonymous site
pNR	proportion of radical (non-conservative) non-synonymous differences per radical non-synonymous site
POMC	pro-opiomelanocortin gene
QTL	quantitative trait locus
rcGH	recombinant carp growth hormone
rcGH I	recombinant carp growth hormone variant I
rcGH II	recombinant carp growth hormone variant II
SRIF14	somatostatin
SRIF	somatostatin
T	testosterone
TRH	thyrotropin-releasing hormone

Trx	thioredoxin
WGD	whole genome duplication

Acknowledgments

This research was supported by a doctoral fellowship from the “Berliner Programm zur Förderung der Chancengleichheit für Frauen in Forschung und Lehre” and technical and financial support from the Leibniz-Institute for Freshwater Ecology and Inland Fisheries, Berlin (IGB).

I give my thanks to Dr. K. Kohlmann (IGB) for providing tissue collection of wild and domesticated common carp populations from Europe, Central and East Asia to study the growth hormone genes. He also provided possibilities to use his laboratory equipment and reagents for carrying out this study. I gratefully appreciate his critical reading of the manuscript and numerous valuable suggestions and critical comments.

I thank Prof. W. Kloas (IGB) for providing possibility to use the method of primary culture of common carp hepatocytes in his laboratory to test recombinant GH activity and for useful discussions.

I wish to express many thanks to Prof. C. Arenz (HU) for possibility to use his laboratory and instruments for production and purification of recombinant common carp growth hormone.

The invaluable support from Prof. A. Elepfandt (HU) and his helpful suggestions and comments on this manuscript have allowed accomplishing this study.

I am thankful to Prof. Bottema (Adelaide University), Dr. Donlon (Galway University), Dr. Herterich (Würzburg University) and Dr. Carlhoff (GE Healthcare) for very useful discussions and suggestions made during my work.

Finally, I like to thank my husband and parents for their help, understanding and encouragement whenever I need it.